

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:35:00 ; Search time 2025.8 Seconds  
(without alignments)  
16520.996 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_1352\_2501

Perfect score: 1150  
Sequence: 1 cgggtaacaagcagcgagac.....attgaggagagcgaagcga 1150

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rod:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_higo\_hum:\*

40: em\_higo\_mus:\*

41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1150	100.0	3442	9	AB008430	AB008430 Homo sapi
2	879.4	76.5	3902	10	BC030329	BC030329 Mus muscu
3	478.8	41.6	2632	10	BC004009	BC004009 Mus muscu
4	335.6	29.2	3719	10	BC009153	BC009153 Mus muscu
5	329.6	28.7	3997	9	AB018336	AB018336 Homo sapi
6	312.8	27.2	2431	10	BC027077	BC027077 Mus muscu
7	241	21.0	106578	9	AL137249	AL137249 Human DNA
8	204.2	17.8	224310	2	AC094777	AC094777 Rattus no
9	199.4	17.3	221502	2	AC122885	AC122885 Mus muscu
10	193.4	16.8	152053	2	AC022669	AC022669 Homo sapi
11	162.4	14.1	393	11	G60059	G60059 SHGC-130885
12	154.6	13.4	426	6	AX261839	AX261839 Sequence
13	150.2	13.1	120911	2	AC130919	AC130919 Rattus no
14	148.6	12.9	190588	2	AC101725	AC101725 Mus muscu
15	148.6	12.9	221502	2	AC122885	AC122885 Mus muscu
16	122.6	10.7	2605	9	BC021301	BC021301 Homo sapi
17	112	9.7	96183	9	AL161896	AL161896 Human DNA
18	112	9.7	152053	2	AC022669	AC022669 Homo sapi
19	111.8	9.7	149417	2	AC118779	AC118779 Rattus no
20	105.8	9.2	42014	2	AC018327	AC018327 Drosophill
21	105.8	9.2	160710	3	AC009537	AC009537 Drosophill
22	105.8	9.2	164713	3	AC007589	AC007589 Drosophill
23	105.8	9.2	303209	3	AE003604	AE003604 Drosophill
24	97	8.4	62537	2	AC101100	AC101100 Mus muscu
25	95.8	8.3	105605	2	AC115336	AC115336 Rattus no
26	95.8	8.3	176630	2	AC109427	AC109427 Rattus no
27	95	8.3	3055	10	BC026860	BC026860 Mus muscu
28	87.8	7.6	134911	9	AC005104	AC005104 Homo sapi
29	83.6	7.3	1296	6	AX411309	AX411309 Sequence
30	83.6	7.3	1821	9	AX002045	AX002045 Homo sapi
31	83.6	7.3	3610	6	AX405929	AX405929 Sequence
32	83.6	7.3	3842	9	AB037783	AB037783 Homo sapi
33	82	7.1	1478	9	BC013319	BC013319 Homo sapi
34	81.4	7.1	3220	5	AF017370	AF017370 Danio rer
35	80.2	7.0	67896	2	AC131316	AC131316 Mus muscu
36	79.6	6.9	133745	2	AC131454	AC131454 Strongylo
37	72.8	6.3	462	6	AX371287	AX371287 Sequence
38	70.8	6.2	138025	9	HS012084	AL096709 Human DNA
39	68	5.9	149417	2	AC118779	AC118779 Rattus no
40	65.8	5.7	2956	10	AF017369	AF017369 Mus muscu
41	64.4	5.6	1896	9	BC032232	BC032232 Homo sapi
42	64.4	5.6	3224	9	BC032429	BC032429 Homo sapi
43	64.4	5.6	4670	9	AK000004	AK000004 Homo sapi
44	62.2	5.4	2170	9	AK097217	AK097217 Homo sapi
45	62.2	5.4	120911	2	AC130919	AC130919 Rattus no

ALIGNMENTS

RESULT 1  
AB008430  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AB008430  
Homo sapiens mRNA for CDEP, complete cds.  
AB008430  
AB008430.1 GI:2766164  
CDEP.  
Homo sapiens embryo cartilage chondrocyte cDNA to mRNA.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Koyano,Y., Kawamoto,T., Shen,M., Yan,W., Noshiro,M., Fujii,K. and  
Kato,Y.  
Molecular cloning and characterization of CDEP, a novel human

3442 bp mRNA linear PRI 13-FEB-1999

protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors  
Biochem. Biophys. Res. Commun. 241 (2), 369-375 (1997)  
98086358

2 (bases 1 to 3442)  
Koyano,Y., Kawamoto,T. and Kato,Y.  
Direct Submission  
Submitted (22-Oct-1997) Takeshi Kawamoto, Hiroshima University  
School of Dentistry, Department of Biochemistry, 1-2-3 Kasumi  
Minami-ku, Hiroshima, Hiroshima 734, Japan  
(E-mail:tkawamo@ipc.hiroshima-u.ac.jp, Tel:082-257-5688,  
Fax:082-257-5629)

## FEATURES

source

Location/Qualifiers  
1..3442  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="chondrocyte"  
/tissue\_type="cartilage"  
/dev\_stage="embryo"  
49..3186  
/function="Rho Guanine Nucleotide Exchange Factor"  
/note="Band 4.1 superfamily"  
/codon\_start=1  
/product="CDEP"  
/protein\_id="BAA24267.1"  
/db\_xref="GI:2766165"  
/translation="MGEIQRTPGSRILGAPENSIGSLERQKPPPTPSKLVSIKI  
QMDDTQAFVQVRAPQKVLDDVNCNLNLNVEGDFGLFPPDHRKRTITWLDLLKPIV  
KOIRRRKVVVFVKKVFFPPDHTQLEELRYLFALQVKODLAOGLRTCNDSAAALI  
SHVQSEIGDFDEALDRHLAKNKYIPQDDALEDRIVFHHNHIGQTPAESDFQLEI  
ARRLMYGRILHPAKDRGTINKLAVANTGILVFOGFTKINAFNWKVKLSFKRFR  
LKLPRDANSVQDVLFLMASRDFCKSFWKICVEHFAFFRLFEPEKPKPVLFSRG  
SSEFSGTQKQVLYVEGKKVKQFPERKSKHSIRLSAQSPHELSEVLEBQSQS  
SLTIFEGEASPGGSGKREPKVPSAGEPSHPSPAPRSPAGNKADGASAPTEE  
EEVVPDQTSQKPPQPPSTGSLTSGPHELSEVNSQGVAPANVTLSPLNSPDTKQ  
ASPLISPLNDQAPRTDEDEGRKREFTDKAYFAKEVSTERTYTKLDELVTISW  
QSTVSKEDAMPALKSLIFPNPEPLHKEFTNFKIEQRLALWEGRSNAQRYDRIQ  
DYLKINQGMKLAHLAKHSEALENGIKSSRRLEFCRDFLOKVCVLPVLTFL  
LPLRLHLYKQVLERLCKHUPPSHADPRCPAALAEJTEWVAQLHGTMIKWNFKL  
HELKDLGIDNLVPGREFIRGLSLKSLGKLQQRMFLLFNDVLLTSGRLTASNQ  
FKVHGOLPIPIESEDWVPHCLTLRGQRQSIIVAASSRSEMKEWEDIQIAD  
LAEKSSPAPELADSPDNKSPDEATAQDEDDLASRTSLEROAPHRGNTMHHV  
CWHRTSVSMVDFSTAVENQSLNLRKFKNSGWQKLWVTFNCFPFYKSHQDNHP  
LASPLLYSLTIPSESENIQDYVFKLHFKSHVYFRASEYTFERWMEVIRSATSS  
ASRPHVLSHKSLSVI"

CDS

Db 1592 CGTGCTGAATGACCAGGCGCTGCCCGGACGACGATGAGGATGAGGCGCGGAGGA 1651  
QY 301 GATTCCTCACTGATTAAGCGTACTTTCATAGCTAAGGAAGTGTCTACCCAGCGGACAT 360  
Db 1652 GATTCCTCACTGATTAAGCGTACTTTCATAGCTAAGGAAGTGTCTACCCAGCGGACAT 1711  
QY 361 ATCTGAAGGATCTCGAAGTTATCAGTTCCTGCTGTTTCAGAGCAGCAGTGACCAAGAGGAGC 420  
Db 1712 ATCTGAAGGATCTCGAAGTTATCAGTTCCTGCTGTTTCAGAGCAGCAGTGACCAAGAGGAGC 1771  
QY 421 CCATCGCGGAAGCAGTGAAGTCTCATATTCGCGAATTTTGAACCTTTGACAAATTTTC 480  
Db 1772 CCATCGCGGAAGCAGTGAAGTCTCATATTCGCGAATTTTGAACCTTTGACAAATTTTC 1831  
QY 481 ATACTAATTTTCTCAAGGAATTTGACCAAGCAGTTCGCTGTTGGGAAGCGCTCAAAATG 540  
Db 1832 ATACTAATTTTCTCAAGGAATTTGACCAAGCAGTTCGCTGTTGGGAAGCGCTCAAAATG 1891  
QY 541 CCCAATCAGAGATTACCAAGAATCGGCGATGTCATGCTGAAGAACAATTCAGGGCATGA 600  
Db 1892 CCCAATCAGAGATTACCAAGAATCGGCGATGTCATGCTGAAGAACAATTCAGGGCATGA 1951  
QY 601 AGCAGCTGGCGGCTCACCTGTGGAAGCAGCAGGAGGCTTTGGAGGCCCTTGGAGAAATGGAA 660  
Db 1952 AGCAGCTGGCGGCTCACCTGTGGAAGCAGCAGGAGGCTTTGGAGGCCCTTGGAGAAATGGAA 2011  
QY 661 TCAAGAGCTCCCGCGGCTGGAGAACTTCTGAGAGACTTTGAGCTGCAGAGAGTGTGTT 720  
Db 2012 TCAAGAGCTCCCGCGGCTGGAGAACTTCTGAGAGACTTTGAGCTGCAGAGAGTGTGTT 2071  
QY 721 ACCTACCGCTCAACACCTTCTCTCGGCGCACTGCACCGGCTCATGCACTACAAGCAGG 780  
Db 2072 ACCTACCGCTCAACACCTTCTCTCGGCGCACTGCACCGGCTCATGCACTACAAGCAGG 2131  
QY 781 TCCTGAGCGGCTGTGCAAAACACCAACCGCGCGACCGCCGACTTCAGGAGCTGCCGAG 840  
Db 2132 TCCTGAGCGGCTGTGCAAAACACCAACCGCGCGACCGCCGACTTCAGGAGCTGCCGAG 2191  
QY 841 CGGCTTTGCGAGAGTACGAGAGTGGTGGCAGAGTCCACAGCTCCACGATGATGATCAAGATG 900  
Db 2192 CGGCTTTGCGAGAGTACGAGAGTGGTGGCAGAGTCCACGATGATGATCAAGATG 2251  
QY 901 AGAATTTCCAGAGCTGCAGAACTCAAGAAAGATTTGATGCGATGACAAATCTTTGTG 960  
Db 2252 AGAATTTCCAGAGCTGCAGAACTCAAGAAAGATTTGATGCGATGACAAATCTTTGTG 2311  
QY 961 TTCGCGGAAGGAGTTCATCGTCTGGGCGAGCTCAGCAAGCTCTCGGGAAGGGGCTCC 1020  
Db 2312 TTCGCGGAAGGAGTTCATCGTCTGGGCGAGCTCAGCAAGCTCTCGGGAAGGGGCTCC 2371  
QY 1021 AGCAGCGCATGTTCTTCCGTTCAACGAGCTGCTGTATACAGCGCGGGGCTCACGG 1080  
Db 2372 AGCAGCGCATGTTCTTCCGTTCAACGAGCTGCTGTATACAGCGCGGGGCTCACGG 2431  
QY 1081 CCTCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGCGATGACGATGAGGAGA 1140  
Db 2432 CCTCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGCGATGACGATGAGGAGA 2491  
QY 1141 GCGAAGACGA 1150  
Db 2492 GCGAAGACGA 2501

RESULT 2

BC030329

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC030329 3902 bp mRNA linear ROD 07-AUG-2002  
Mus musculus, clone IMAGE:5376197, mRNA, partial cds.  
BC030329  
BC030329.1 GI:20987935  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

polya\_site

/note="50 a nucleotides"

BASE COUNT 864 a 952 c 927 g 699 t

ORIGIN

Query Match 100.0%; Score 1150; DB 9; Length 3442;  
Best Local Similarity 100.0%; Pred. No. 6e-266;  
Matches 1150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGTAACAGCAGGCGGAGCGCGCTCGCGCGCCCGCAGGAGGAGGAGGAGGTCG 60

Db 1352 CGGGTAACAGCAGGCGGAGCGCGCTCGCGCGCCCGCAGGAGGAGGAGGAGGTCG 1411

QY 61 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCCCGCAGGCAAGCAGGCTCCCTGA 120

Db 1412 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCCCGCAGGCAAGCAGGCTCCCTGA 1471

QY 121 CTGGCAGTCTCAGCTTCCGAGCTCTCTGAACTCGCAGGCGGAGTGGCCCTGCCA 180

Db 1472 CTGGCAGTCTCAGCTTCCGAGCTCTCTGAACTCGCAGGCGGAGTGGCCCTGCCA 1531

QY 181 ACGTGACCTTGTCTCCCACTCAGCGCCGACACCAAGCAGGCGCTCTCCCTGATCAGCC 240

Db 1532 ACGTGACCTTGTCTCCCACTCAGCGCCGACACCAAGCAGGCGCTCTCCCTGATCAGCC 1591

QY 241 CGTGTGTAATGACAGGCGTCCCGCGGACGAGCAGTGAAGGAGGCGCGGAGGA 300



Db 1747 GACTGACGCATCTATCATGTTAAAGTCCAGGACGATCCCACTCTATGCGATGACGA 1806  
QY 1132 TTGAGGAGAGCGAAGACGA 1150  
Db 1807 TCAGGAGAGTGGAGGA 1825

RESULT 3  
BC004009 2632 bp mRNA linear ROD 07-AUG-2002  
LOCUS Mus musculus, clone IMAGE:3493093, mRNA, partial cds.  
DEFINITION BC004009  
ACCESSION BC004009.1 GI:13278387  
VERSION  
KEYWORDS house mouse.  
SOURCE  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2632)  
Direct Submission  
Submitted (28-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 8 Row: P Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.

FEATURES  
Location/Qualifiers  
1..2632  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="CZECH II"  
/clone="IMAGE:3493093"  
/tissue\_type="Mammary tumor metastasized to lung. Tumor  
arose spontaneously from a senescent normal mammary  
(clonal) outgrowth infected with the virus MMV."  
/clone\_lib="NCI\_CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
<1..1263  
/codon\_start=1  
/product="Unknown (protein for IMAGE:3493093)"  
/protein\_id="AAH04009.1"  
/db\_xref="GI:13278388"  
/translation="VMLNKIQMKHLAAHLWKHSEALEATSIKSRRLHFRCDFE  
LQKCYLPNTFLPLRLHMYKHVLELCKHPPNHADFCRAALAEITEMVAQL  
HGTHLKMENFQKLEHLKDLGLDNLVPGREFIRGLSKLSKGLQORFFFNVDV  
LLYTRSGTANQKRVHGOGLPYGMTIEESEEWGVPCHLIRGQRSDIIVAASRSE  
MEKMDTQAMIDLAENSGNPTPELLASPPDKSPDPAADSEDDLSAASRTSL  
RQAPHRGNTVHVHWRSTSVMDVFSIAVENQLSGNLLRFRKNSGWQKWLWVFTNF  
CLFFYKHQDSHPLASLPGLGYSLTIPSENIHKDYVFKLHFKSHVYFFRAESEYTF  
ERWMEVIRSATSSASRAHLHSHKESHLY"

BASE COUNT 632 a 720 c 671 g 609 t

ORIGIN

Query Match 41.6%; Score 478.8; DB 10; Length 2632;

Best Local Similarity 89.3%; Pred. No. 1.e-104;  
Matches 516; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 573 GTCATGCTGAAGAACAATTCAGGGCATGAAGACACCTGGCGGCTCACCTGTGGGAAGACACAGC 632  
Db 1 GTCATGCTGAAGAACAATTCAGGGCATGAAGACACCTGGCGGCTCACCTGTGGGAAGACACAGT 60  
QY 633 GAGGCTTTGGAGGCGCTTGAGAAATGAATCAAGAGCTCCCGGCGGTGGAGAACTTCTCGC 692  
Db 61 GAGGCGCTTGAGGCGCTTGAGAAATGAATCAAGAGCTCCCGGCGGTGGAGAACTTCTCGC 120  
QY 693 AGAGACTTTGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCTCTCTGGGGCCA 752  
Db 121 CGAGACTTCGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCTCTCTGGGGCCA 180  
QY 753 CTGACCGGCTCATGCACTACAGAGGTCTCTGGAGCGGCTGTGCAAAACACACCGCGCG 812  
Db 181 CTGACCGGCTCATGCACTATTAAGCATGTCTCTGGAGAGGCTGTGCAAGACACACCAACCA 240  
QY 813 AGCCAGCGGCTTCAGGAGCTCCGAGCGCTTTGGCAGAGATCAGGAGATGTTGGCA 872  
Db 241 AACCGAGCGGCTTCAGGAGCTGCGAGAGCTGCGGTGGCGGAGATCACAGAGATGTTGGCC 300  
QY 873 CAGCTCCAGCGGTACGATGATCAAGATGGAGAATTTCCAGAAAGCTGCACGAAGCTCAAGAAA 932  
Db 301 CAGCTGCAGCGGCGCATGATCAAGATGGAGAATTTCCAGAAAGCTGCATGAGCTCAAGAAA 360  
QY 933 GATTGATTGGCATTTGACAAATCTTGTGGTCCGGAAGGAGTTTCATCGTCTGGGCAGC 992  
Db 361 GATCTGATCGGCATTTGACAAATCTTGTGATCCAGGAAGGAGTTTCATCGCTGGGCAGC 420  
QY 993 CTCACAAGCTCTCGGGAAGGCGCTCCAGCAGCGCATGTTCTCTCTTCAACACGTC 1052  
Db 421 CTCACAAGCTCTCGGGAAGGCGCTTCAGCAGCGCATGTTCTCTCTTCAACAGTC 480  
QY 1053 CTGCTATACAGCAGCGGCGGCTGACGGCTCCAATCAGTTTAAAGTCCACGGGAGGTC 1112  
Db 481 TTGCTGTATACAGCGGCGGCTGACAGCATCTAATCAGTTTAAAGTCCACGGAGGTC 540  
QY 1113 CGCTCTATGCGATGACAGTTGAGGAGCGGAGAGCA 1150  
Db 541 CCACCTCTATGCGATGACGATCGAGGAGAGTGGAGGA 578

RESULT 4  
BC009153  
LOCUS BC009153 3719 bp mRNA linear ROD 07-AUG-2002  
DEFINITION Mus musculus, Similar to KIAA0793 gene product, clone MGC:6304  
IMAGE:2655209, mRNA, complete cds.  
ACCESSION BC009153  
VERSION BC009153.1 GI:14318718  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 3719)  
Strausberg, R.  
Direct Submission  
TITLE  
JOURNAL

REMARK  
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>



Contact: amg@bcm.tmc.edu  
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 5 Row: m Column: 20.

#### FEATURES

Location/Qualifiers  
1..3719  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB-N-3"  
/clone="MGC:6304 IMAGE:2655209"  
/tissue\_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month  
old mouse. Taken by biopsy."  
/clone\_lib="NCI CGAP\_Mam2"  
/lab\_host="DH105"  
/note="Vector: pCMV-SPORT6"  
99..3296  
CDS  
/codon\_start=1  
/product="Similar to KIAA0793 gene product"  
/protein\_id="AAH09153.1"  
/db\_xref="GI:14318719"  
/translation="MGEIEGTYRALPTSGRLGGQTAIGVSTLEPQSLSPRMQEKHM  
RIRVLLDSVFLVLEIEPKDCQVLLTQWKKHLNTECDYFGLKFNKVSQYIWLPEPM  
KPIIRQVRKPNVLRVAVKFPDPDQLOEEYTRYLPALQRLLERLTCTANTA  
ALLSHLQSEIGDYDETDREHLKANEVLPNOEKSLEKILDFHORHTGOTPAESDFQ  
VLEIARKLEMGIRPHMSDREGTKINLAVSHMGVLVFGQTKINTFNKVRKSLFV  
RKRLKLHPEVHGYPQDTLEFLGSDCKNFWKICVEYHFFSLSDQPKPKAVF  
FSGSFRYQKQVLDVYKDGKRIPIYERHRSKRTLSHALTDLPKQSVSFTD  
GLRTSASSANVSYPSPSSLPGLNPKDSSSLVDPOAPVTKSTAARSQGPS  
SSDGSTQSLHLPVPLRPGFSGMSPSPSLKSHLSLCPLOALALSTAEQGS  
PVLSPVLSAGTARMNDQEEOKHMPDEAYEIAKEILATERVYLKOLEVTVWFRS  
VLJKEAPMALMALLFNSIDPVEYFHRGFLHEVEQRLALWEGSPSAHLKGBQIRIG  
ILRNMRQKFTSFORHDEVTELEKATKCKKLEAVYEFELQKVCYLPENLFL  
KPVRLHYRLLSLCAHYSPGHRDYADCHEALKAITEVTELEQOQSLRLLENLQKLT  
EQRDLVGENIAPGREFIREGLKHLTKGLQRMFFLFSDMLLYTSKQDTSASHF  
RIRGLPLRGLMEVESENEVSLHCTIYAAQKTIIVAASTRLEKEKMDLNAIQ  
ACTIGDSPVLLGAVNYLTPRSDSEVLEESDGRNRSLEGNHOKRANTMHVCW  
YRNTSVSRDASHAAVENOLSGYLLRKFKNNGWKLWVFTNFCFLFFVYKTHODDYPLA  
SLPLGLYSVSLPREADSHKDYFKLQPKSHVYFPRASKYTFERMMOVIKRASSSPG  
RPPSTQCSHSPGLEAEIREKEACPPSLDKNL"

BASE COUNT 997 a 979 c 929 g 814 t  
ORIGIN  
Query Match 29.2%; Score 335.6; DB 10; Length 3719;  
Best Local Similarity 60.1%; Pred. No. 5.2e-70;  
Matches 576; Conservative 0; Mismatches 379; Indels 3; Gaps 1;  
QY 196 CCAACCTGAGCCCCGACACCAAGCAGCGCTCTCCCTGTGATCAGCGCGCTGCTGAATGACC 255  
Db 1591 CCGCGCTGAGTACAGCTGAGCAGGGTGATCCCTCTGTCTCAGCGCTGTCTCAGTGGCG 1650  
QY 256 AGGCTGCCCCCGAGGAGGATGAGTGGCGCGGAGGAAGATGCCCAACTGATA 315  
Db 1651 CTGGTACAGCGAGGATGGACACCAAGAGAGAGACAGAACACAGACATGCCAAGAGATG 1710  
QY 316 AAGCGTACTTCATAGTAAAGAGTCTCTACCAACCGAGCGAAATATCTCAAGGATCTCG 375  
Db 1711 AGCCCTATTTTCATAGCAAGAGATCTCTCGCTACAGCAAGCACTCTGAAGGATTTAG 1770  
QY 376 AAGTTATCACTTCCTGTTTCAGACACAGTGGAGAAAGAGCGCATGCCCGGAAGCAC 435  
Db 1771 AAGTTATCACTGTGTTCCGAGCGCTGTGATCAAGGAGGAGCGCTGCGTGCAGCCC 1830  
QY 436 TGAAGAGTCTCATATTCGGAATTTTGAACCTTTTGACAAATTTTCACTAATTTTCTCA 495  
Db 1831 TGATGGCCCTGCTTTCTCCAACTGATTCACGACTCTACGAGTTCCACAGAGGCTTCTCTC 1890  
QY 496 AGGAATTTGAGCAACACACTTGCCTGTGGAGCGCGCTCAATAGCCCA---AATCAGAG 552  
Db 1891 ACAGGTGGAACAGAGGCTGGCACTCTGGAGAGGCGCTCCAGTGCCTTAAAGAGTG 1950

QY 553 ATTACCAAGAATCGGCGATGTCTATGCTGAAGAACAATTCAGGGCATGAAGCACTTGGCGG 612  
Db 1951 ATCACCAGCAATCGGGACATCTCTCTAGGAACATGCGTCAGTTAAAGGAATTTACTA 2010  
QY 613 CTCACCTGTGGAAACACAGACGAGCGCTTGGAGCCCTTGGAGAAATGAATCAAGAGTCTCC 672  
Db 2011 GCTACTTCCAAAGACACGATGAGGTCTTAACAGAACTGGAAAAGGCCACAAAACACTGTA 2070  
QY 673 GCGCGGTGGAGAACTTCTCAGAGACTTTGAGCTGAGAGAGTGTGTACCTACCTACCGCTCA 732  
Db 2071 AAAAGCTGGAGGAGCTTACAAAGAGTTGAGCTCCAAAAGGTTCTGCTACTCTCCCTCTCA 2130  
QY 733 ACACCTTCTCTCTCGCGGCGCTCATGACCGGCTCATGACATAACAAGCAGGTCTCTGGAGCGG 792  
Db 2131 ACACATTTCTGTCTGAAGCCGCTCCAGAGGCTAGTCCACTACCGCTCTGCTGCTGAGCGGC 2190  
QY 793 TGTGCAAAACACACCGCGGAGCCAGCGGACTTTCAGGAGACTCCCGAGCGGCTTTTGGCAG 852  
Db 2191 TGTGTGCTCACTACTCTCTCTGGGCGGCGACTATGCGGACTGCCATGAGGCACATGAAGG 2250  
QY 853 AGATCAGGAGATGTTGGGACACAGCTCCACGGTACGATGATCAAGATGGAGAATTTCCAGA 912  
Db 2251 CCATCAGAGAAGTGACCCAGCGCTCCAGCAAAAGCCCTTACCCGCGCTGGAAAACCTACAGA 2310  
QY 913 AGCTGCACGAACCTCAAAAAGATTTGATTGGCATTGACAAATCTTGTGGTTCCGGGAAGGG 972  
Db 2311 AATTGAGGAGCTACAGCGAGACTGTCGGTGTAGAAAACCTCATTGCTCTCTGGGAGGG 2370  
QY 973 AGTTCATCGTCTGGGACCGCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGT 1032  
Db 2371 AGTTTATCGTGGGCTGCTGCATCAAGCTCAACAAAGAGGCGCTGCAACACAGGATGT 2430  
QY 1033 TCTTCTGTTTCACAGCGTCTCTATACAGAGCGCGGGGCTGACGGGCTTCCAAATCAGT 1092  
Db 2431 TTTTCTGTTCTCAGATATGTTTACTGTATACAAGCAAAAAGTGTACAGAGGACCGACGTCAT 2490  
QY 1093 TTAAGTCCACGGGCGAGCTCCCTCTCTATGGCATGACGATTGAGGAGAGGAAGACGA 1150  
Db 2491 TCCGGATCGTGGCTTCCTTCCACTCCGCGCATCGTGTAGAGAAAGTGTAGAAATGA 2548  
RESULT 5  
AB018336  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
JOURNAL  
FEATURES  
source  
AB018336  
Homo sapiens mRNA for KIAA0793 protein, complete cds.  
AB018336  
AB018336.1 GI:3882306  
Homo sapiens adult male brain cDNA to mRNA, clone\_lib:pBluescriptII  
SK plus clone:hk05692.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,  
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.  
Prediction of the coding sequences of unidentified human genes. XI.  
The complete sequences of 100 new cDNA clones from brain which code  
for large proteins in vitro  
DNA Res. 5 (5), 277-286 (1998)  
2 (bases 1 to 3997)  
Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.  
Direct Submission  
Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute,  
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba  
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,  
Fax:+81-438-52-3914)  
Location/Qualifiers  
1..3997  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

```

/clone="hk05692"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
1..3997
/gene="KIAA0793"
118..3282
/gene="KIAA0793"
/codon_start=1
/product="KIAA0793 protein"
/protein_id="BAA34513.1"
/db_xref="GI:3882307"

```

/translation="MGEIEGTYRVLQTAGNRLCAQTPVGSTLLEPQOTLLPRMQEKHL  
 HLRVLTONTMEIFLDEPKCGQVLTQVWKNRLNVECYDFGMEFQNTQSYWIMLEPM  
 KPIIROIRRNPNVFLERKAFPPDQGOQEETRYLFALOKRLDLLEERLQWICADPTTA  
 ALATSHLLQSEIGDITBMSDREKHTNXYLPGQOGLKEIKLEHQHNFQVQTPAESDFQ  
 VLEATRLMEWIGRFGHMSAREGHTKQLAVSHMGVLVFOGTTKINTFNNSKYVRKLSFK  
 RKRFILKHPHEVHGPODYTLFEFLGSRDECKNFWKTCVEYHUPFRLLDOPKPKAKAVE  
 FSRGSPRYSGRTQKVLVDYFKDSGNKRIPYERHSKTHTSVRALYADLPKQSISEPE  
 GLRTPASPGRATYKSLPSTLVPFSGPEFKDSSSLSTQPVSVSPAPRAERSGAVAF  
 GGPVPSDAGGMPALGPQGLSTKSPQSPSSRKPLSLSPAFQVPLCPGAPQGSPL  
 LSPVSDAGGAGMDECEPRHKRYPADAEYIVKEILATERTYLDKLEVITVWFRSAAV  
 KEDAMPATLMTLFLSDNIDITFPHRGFLREVEORLALWEGPSKAHKGSHGRIOTLL  
 RNMOLAKFTYSYFORHDEVLTEKATCKRCKLEAVYKFELOKQVCYLLPLNTFLFKPI  
 ORLLHYLLRLRLGGHYSPGHDPYACHDAKKAITEVTTLOHILIRLENLOKLTLO  
 RDLVGLENLALPGREFIRECECLKLTKKQLOQRMFLFSDMLLYLSKGVAGVSHFRIR  
 GLLPLOQMLVEESDNWSVPHCFITVAAQKTIIVAASTRLEKEKMWLDNSAIQAARK  
 GGDTPALPACVNTOPRPSLNVEJSEDEDDARGVRSLEGGHQRHANTHMIVCQV  
 RNTYSVRADTHSAAVENLOSPSLKFNKNSHGQWLKWVFTNCFLYTKTHDDYPLAS  
 LPLUGYSVSIPEADGTHKDYVEFKLQYKSHVYFFRAESKYTFERNNEVIQGASSSARG  
 APSTVPOGPOPSSEICGMVRGKEP"

BASE COUNT	998 a	1132 c	1068 q	799 t
------------	-------	--------	--------	-------

Query Match	28.7%	Score 329.6;	DB 9;	Length 3997;
Best Local Similarity	60.6%;	Pred. No. 1.4e-68;		
Matches 577;	Conservative 0;	Mismatches 369;	Indels 6;	Gaps 2;
QY	202	TGAGCCCGACACCAAGCAGGCGCTTCCTCTGATCAGCGCGCTGCTGAATCACCAGGCGCT	261	
DB	1610	TGGGCCCAGCTGACAGGGCTCATCCCACTCTCTGAGCCCTGTCTCTAGTGATCTGGCG	1669	
QY	262	GCCCCCGACGAGCGATGAGGATGAGGCGCGGAGGAGATGCCAACTGATAAAGCGT	321	
DB	1670	GAGCCGGGATGGAGTGGGAG---GAGGCCACACACAAGCGCGTCCCTGCACACAGGCGCT	1726	
QY	322	ACTTCATAGCTAAGGAAGTGTCTACCCAGCGAGCAACATATCTCAAGGATCTCAAGTTA	381	
DB	1727	ACTTCATAGTCAAGAGATTCTCGCTACAGAACCAACATACCTCAGGATTTGAAGTTA	1786	
QY	382	TCACTTCGTGTTTTCAGAGCACAGTGAAGAGAGAGCGCATCCGCGAAGCACTGAAAA	441	
DB	1787	TTACCGTGTGTTTCCGACGCGACTGTGTGAAGGAGGAGCGCCATGCCCTCGGACTCTGATGA	1846	
QY	442	GTCTCATATTTCCCGAATTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAA	501	
DB	1847	CGTGTCTTCTTCCCAATCATGATCCCATCTATGATGTCCACAGAGGCTTCCTGCGCGAGG	1906	
QY	502	TTGAGCAACGACTTGCCTGTGGCAAGCGCCCTCAAATGCCAAATCAGAG---ATTACC	558	
DB	1907	TGGAGCAGAGCTGGCACTCTGGGAAGGCGCTCCAAAGCCCAACAAAGGCAGTCATC	1966	
QY	559	AAAGAATCGGGATGTCTATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGCTCAC	618	
DB	1967	AACCAATCGGGAGATCTGTCTCAGGAACATGCGCCAGTTAAAGGAGTTTACCAGCTACT	2026	
QY	619	TGTGGAAGCAGCAGGCGCTTTGGAGCGCCCTGGAGAGATGGAATCAAGAGCTCCCGCGGC	678	
DB	2027	TCCAAAGCATGACGAGGTCTTAACAGACTGGGAAAGGCTACCAAAGCCTGTGAAGT	2086	
QY	679	TGGAGAATCTCTCGAGAGACTTTGAGCTGCAAGAGGTGTGTACTTACCGCTCAACACT	738	

Db	2087	TGGAGGCGAGTGTACAAGGAGTTTGAGCTGCAGAGGTTCTGCTACTTGTGCGCTCTCAACACGT	2146
Qy	739	TCCCTCCGCGCCACTGCACCGGCTCATATGACTACAAGCAGGTCCTGGAGCGGCTGTGCA	798
Db	2147	TCCTGCTGAAGCCATCCAGGGCTGCTGCACCTACCCCTGCTGCTGCGCGGCTATGCG	2206
Qy	799	AACACCACCGCGAGCCACCGGACTTCAGGAGCTGCCAGCGGCTTTGGCAGAGATCA	858
Db	2207	GACATTACAGCCCGGCGACCATGACTACGTGACTGCCATGACGCCCTCAAGGCCATCA	2266
Qy	859	CGGAGATGGTGGACACACTCCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGC	918
Db	2267	CAGAGTGCACACACACATACGACATTCATCCGGCTGGAGAACCTCAGAAGCTAA	2326
Qy	919	ACGAACCTCAAGAAAGATTTGATTGGCAATTCACATCTTGTTGGTTCGGGGAGGAGGTCA	978
Db	2327	CGGAGCTGCAGCGGACTGTTGGGCATAGAGAACCTCATTTGCTTCGCGGAGGAGTTCA	2386
Qy	979	TCCGCTCTGGGAGCGCTCAGCAAGCTCTCGGGGAGGCGGCTCCAGCAGCGCATGTTCTTC	1038
Db	2387	TCCGTGAGGGCTGCTTCAACAAGCTCAACAAGAAGGCGCTGCAGCAGAGGATGTTTTC	2446
Qy	1039	TGTTCAAGGAGGCTCTGCTATACAGAGCGGGGCTGACGGGCTTCCAATCAGTTTAAAG	1098
Db	2447	TGTTCTCAGATATGTTGCTGTACACAAGAAAGAGTTCAGAGGACCGACCTCCCGGA	2506
Qy	1099	TCCACGGGAGCTCCCGCTCTATGCGATGACGATTGAGGAGAGCGAAGACGA	1150
Db	2507	TCCGGGCGCTTCTCCCGCTCAAGGCATGCTGGTGGAGAAAGTGATAACGA	2558
RESULT 6			
LOCUS	BC027077		
DEFINITION	Mus musculus, clone IMAGE:5010682, mRNA	linear	ROD 07-AUG-2002
ACCESSION	BC027077		
VERSION	BC027077.1	GI:20071584	
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2431)		
TITLE	Strausberg, R.		
JOURNAL	Direct Submission		
	Submitted (04-APR-2002) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk		
	Email: <a href="mailto:cgabbs-femail.nih.gov">cgabbs-femail.nih.gov</a>		
	Tissue Procurement: Gilbert Smith, Ph.D.		
	CDNA Library Preparation: Life Technologies, Inc.		
	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Baylor College of Medicine Human Genome		
	Sequencing Center		
	Center code: BCM-HGSC		
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a>		
	Contact: <a href="mailto:amg@bcm.tmc.edu">amg@bcm.tmc.edu</a>		
	Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,		
	Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,		
	Richards, S., Gibbs, R.A.		
	Clone distribution: MGC clone distribution information can be found		
	through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Series: IRAC Plate: 45 Row: c Column: 12		
	This clone was selected for full length sequencing because it		
	passed the following selection criteria: Hexamer frequency ORF		
	analysis.		
FEATURES			
source	Location/Qualifiers		
	1..2431		
	/organism="Mus musculus"		

```
/db_xref="taxon:10090"
/map="FVB/N-3"
/clone="IMAGE:5010682"
/tissue_type="mammary tumor. MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<!.1065
CDS
/codon_start=1
/product="Unknown (protein for IMAGE:5010682)"
/protein_id="AAH27077.1"
/db_xref="GI:20071585"
/translation="YKHLERLCKHPPNHADFRCAALAEITEMVAQLHGTMIKME
NFQKLEKLDIGDNLVTPGRFELKLSKGLQORMFFLNDVLLYTSRGL
TASNOFKVHGLPLGVTIEESEEWGPHCLTLRGORSIIAASRSRSEKMWEDI
QMAIDLAKNSGTPPELLASSPPDNKSPDEATAADQSEDDLSASRTSLERQAPHRGN
TWVHCWHRSTSVSWDESIAVENQLSRKFNKNSGWKLWVFNFCUFLFFYKSH
QDSHPLASLPLGLSLTIPSESENTHDKYVFLHFKSHVYFRASEYTFERWMEVIR
SATSSASRAHILSHRESHLY"
BASE COUNT      581 a   659 c   614 g   577 t
ORIGIN
Query Match      27.2%; Score 312.8; DB 10; Length 2431;
Best Local Similarity 88.9%; Pred. No. 1.5e-64;
Matches 338; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 771 TACAAGCAGGTCTCGAGCGGCTGTGCAACACACCACCGCCGAGCCAGCGACTTCAGG 830
Db 1 TATAAGCATGTCTCGAGAGGGCTGTGCAAGCAGCCACCCACCAACACCGCGACTTCAGG 60
QY 831 GACTGCGGAGCCGCTTTGGCAGAGATCAGCGAGATGGTGCGACAGCTCCACAGGTACGATG 890
Db 61 GACTGAGAGCTCGCTGGCGGAGATCAGAGATGGTGCGACAGCTGATGTACCATG 120
QY 891 ATCAAGATGAGCAATTTCCAGAACTGCGACGAACTCAAGAAAGATTGATGGCATTTGAC 950
Db 121 ATCAAGATGAGAACTTCCAGAACTGCGATGAGCTCAAGAAAGATCTGATCGGCATTGAC 180
QY 951 AATCTTGTGTCTCGGAGAGGAGTTTCATCCGCTGCGGAGCCTCAGCAAGCTCTCGGG 1010
Db 181 AATCTTGTGACCCAGAGAGGAGTTTCATCCGCTGCGGAGCCTCAGCAAGCTCTCGGG 240
QY 1011 AAGGGCTCAGCAGCGCATGTTCTTCTGTTCAACGAGCTCTGCTATACAGAGCCGG 1070
Db 241 AAGGGCTCAGCAGCGCATGTTCTTCTGTTCAACGAGCTCTGCTGTTATACAGCCGG 300
QY 1071 GGCTGACGCGCTCCATCAGTTTAAGTCCAGCGGCGCTCCGCTCTATGGCATGACG 1130
Db 301 GGACTGACGCGCATCTAATCAGTTTAAAGTCCAGCGAGCTCCCACTCTATGGCATGACG 360
QY 1131 ATTGAGGAGAGCGAAGACGA 1150
Db 361 ATCAGGAGAGTACGAGGA 380
RESULT 7
AL137249
LOCUS      106578 bp DNA linear PRI 12-FEB-2002
DEFINITION Human DNA sequence from clone RP11-111L24 on chromosome
13q31.3-32.3, complete sequence.
ACCESSION AL137249
VERSION   AL137249.29 GI:18655955
KEYWORDS  HTG.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106578)
AUTHORS   Garner,P.
TITLE     Direct Submission
JOURNAL   Submitted (12-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
```

COMMENT

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Feb 13, 2002 this sequence version replaced gi:18121451.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em:, EMBL; Sw:,  
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 13, constructed by the Sanger Centre Chromosome 13  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr13  
RP11-111L24 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone  
RP11-111L24 It may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true left end of clone RP11-295B17 is at 104579 in this  
sequence. The true right end of clone RP11-261P24 is at 2000 in  
this sequence

FEATURES

source  
1..106578  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="q31.3-32.3"  
/clone="RP11-111L24"  
/clone\_lib="RPCI-11.1"  
79533  
misc\_feature  
/note="Tandem repeat. Forced join. Gap size estimated to  
be approximately 800bp by restriction digest data"  
BASE COUNT 28669 a 25979 c 25140 g 26790 t  
ORIGIN

Query Match 21.0%; Score 241; DB 9; Length 106578;  
Best Local Similarity 100.0%; Pred. No. 3.5e-47;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 600 AAGCACCCTGGCGGCTCACCTGTGGAAGCAGCAGCGGCGCTTGGAGGCCCTCGAGATGGA 659  
Db 37298 AAGCACCCTGGCGGCTCACCTGTGGAAGCAGCAGCGGCGCTTGGAGGCCCTCGAGATGGA 37357  
QY 660 ATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTTCAGCTGCAGAAAGGTGTGT 719  
Db 37358 ATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTTCAGCTGCAGAAAGGTGTGT 37417  
QY 720 TACCTACCGCTCAACACCTTTCCTTCCTGCGGCCACTGCACCGGCTCATGCACTACAAGCAG 779  
Db 37418 TACCTACCGCTCAACACCTTTCCTTCCTGCGGCCACTGCACCGGCTCATGCACTACAAGCAG 37477  
QY 780 GTCCTGGAGCGGTGTGCAACACACACCGCGGAGCCGCGGAGCTTCAGGGAGCTCCCGA 839  
Db 37478 GTCCTGGAGCGGTGTGCAACACACACCGCGGAGCCGCGGAGCTTCAGGGAGCTCCCGA 37537  
QY 840 G 840  
Db 37538 G 37538

## RESULT 8

AC094777

LOCUS

DEFINITION

AC094777 224310 bp DNA linear HTG 24-AUG-2002  
 Rattus norvegicus clone CH230-4N8, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 65 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC094777.6 GI:22164997  
 HTG: HTGS\_PHASE1.  
 Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

## REFERENCE

AUTHORS

1 (bases 1 to 224310)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Bubay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorusu, L., Loul, L., Loul, L., Loul, L., Loul, L., Loul, L.,  
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mahoney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,  
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,  
 Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,  
 Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwuonu, G.,  
 Olarunpogoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,  
 Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A.,  
 Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,  
 Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,  
 Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,  
 Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,  
 Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,  
 Simon, I., Sitter, C. D., Smajda, D., Sneed, A., Sodergren, E.,  
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,  
 Svatek, A., Taber, P., Taylor, K., Taylor, T., Thomas, N., Thomas, S.,  
 Tinney, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,  
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,  
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R.,  
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,  
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,  
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,  
 Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

## TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

AUTHORS

TITLE

Direct Submission

## JOURNAL

## COMMENT

Submitted (24-AUG-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Aug 9, 2002 this sequence version replaced gi:22094284.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBIZ

Center clone name: CH230-4N8

----- Summary Statistics

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 172428 bases at least Q40

Consensus quality: 179331 bases at least Q30

Consensus quality: 184096 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 65 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1011: contig of 1011 bp in length

\* 1012 1111: gap of unknown length

\* 1112 2458: contig of 1347 bp in length

\* 2459 2558: gap of unknown length

\* 2559 3836: contig of 1278 bp in length

\* 3837 3936: gap of unknown length

\* 3937 5467: contig of 1531 bp in length

\* 5468 5567: gap of unknown length

\* 5568 6863: contig of 1296 bp in length

\* 6864 6964: gap of unknown length

\* 6964 8013: contig of 1050 bp in length

\* 8014 8113: gap of unknown length

\* 8114 9431: contig of 1318 bp in length

\* 9432 9531: gap of unknown length

\* 9532 10970: contig of 1439 bp in length

\* 10971 11070: gap of unknown length

\* 11071 12178: contig of 1108 bp in length

\* 12179 12278: gap of unknown length

\* 12279 13604: contig of 1326 bp in length

\* 13605 13704: gap of unknown length

\* 13705 15160: contig of 1456 bp in length

\* 15161 15260: gap of unknown length

\* 15261 16520: contig of 1260 bp in length

\* 16521 16620: gap of unknown length

\* 16621 18279: contig of 1659 bp in length

\* 18280 18379: gap of unknown length

\* 18380 19562: contig of 1183 bp in length

\* 19563 19662: gap of unknown length

\* 19663 20728: contig of 1066 bp in length

\* 20729 20828: gap of unknown length

\* 20829 22059: contig of 1231 bp in length

\* 22060 22159: gap of unknown length

\* 22160 24977: contig of 2818 bp in length

\* 24978 25077: gap of unknown length

\* 25078 26432: contig of 1355 bp in length

\* 26433 26532: gap of unknown length

\* 26533 27664: contig of 1132 bp in length

\* 27665 27764: gap of unknown length

\* 27765 29834: contig of 2070 bp in length

\* 29835 32076: gap of unknown length

\* 32077 32176: contig of 2142 bp in length

\* 32177 33953: contig of 1777 bp in length

\* 33954 34053: gap of unknown length



```

* 42440 71538: contig of 29099 bp in length
* 71539 71638: gap of unknown length
* 71639 99587: contig of 27949 bp in length
* 99588 99687: gap of unknown length
* 99688 134048: contig of 34361 bp in length
* 134049 134148: gap of unknown length
* 134149 221349: contig of 87201 bp in length
* 221350 221449: gap of unknown length
* 221450 221502: contig of 53 bp in length.
FEATURES
    source
        1..221502
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="UNK"
            /clone="RP23-90H14"
            1..8441
                /note="assembly_name:Contig59"
            8542..27277
                /note="assembly_name:Contig60"
            27378..42339
                /note="assembly_name:Contig61"
            42440..71538
                /note="assembly_name:Contig62"
            71639..99587
                /note="assembly_name:Contig63"
            99688..134048
                /note="assembly_name:Contig64"
            134149..221349
                /note="assembly_name:Contig65"
            221450..221502
                /note="assembly_name:Contig19"
        61405 a 54522 c 50450 g 54373 t 752 others
    ORIGIN

```

```

Query Match 17.3%; Score 199.4; DB 2: Length 221502;
Best Local Similarity 89.2%; Pred. No. 3.6e-37;
Matches 215; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 600 AAGCAGCTGGCGGCTCAGCTGTGGAGACAGCAGGAGCGCTGGAGCGCTGGAGATGGA 659
|||||
Db 74064 AAGCAGCTGGCGGCGGCTCAGCTGTGGAGACAGCAGGAGCGCTGGAGCGCTGGAGACCTCC 74123
|||||
QY 660 ATCAAGAGCTCCGCGGCTGGAGAACTTCTGAGAGACTTTGAGCTGCAGAGGTGTGT 719
|||||
Db 74124 ATCAAGGCTCGCGGCGCTGGAGAACTTCTGCGAGACTTCTGAGCTGCAGAGGTGTGT 74183
|||||
QY 720 TACCTACCGCTAACACCTTCTCTCGGGCGACCTGCACCGGCTATGCACTACAGCAG 779
|||||
Db 74184 TACCTCGCGCTCAACACCTTCTCTCGGGCGACCTGCACCGGCTCATGCACTATAGCAT 74243
|||||
QY 780 GTCTGGAGCGGCTGTGCAACACACCGCGGAGCGGAGCGGAGCTTCAGGAGCTGCCGA 839
|||||
Db 74244 GTCTGGAGCGGCTGTGCAAGAGCACCACCCACCAACCGGAGCTTCAGGAGCTGCAGA 74303
|||||
QY 840 G 840
Db 74304 G 74304

```

```

RESULT 10
AC022669
LOCUS
DEFINITION Homo sapiens chromosome 13 clone RP11-96B23 map 13, HTG 24-AUG-2002
SEQUENCE, 33 unordered pieces.
ACCESSION AC022669
VERSION AC022669.4 GI:9966291
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 152053)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

# TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens chromosome 13, clone RP11-96B23  
Unpublished  
2 (bases 1 to 152053)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivart,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severly,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

# TITLE JOURNAL REFERENCE AUTHORS

Direct Submission  
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 152053)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,  
Canarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,  
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,  
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
Iliev,I., Johnson,R., Jones,C., Karatas,A., Lander,E.,  
Landares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,A., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

# TITLE JOURNAL COMMENT

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 4, 2000 this sequence version replaced gi:7229803.  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: 96\_B\_23  
Center clone name: L4938  
----- Summary Statistics  
Sequencing vector: M13; M7815; 93% of reads  
Sequencing vector: Plasmid; n/a; 0.0% of reads  
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of  
reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 136451 bases at least Q40  
Consensus quality: 143612 bases at least Q30  
Consensus quality: 146365 bases at least Q20  
Insert size: 158000; agarose-fp  
Insert size: 148853; sum-of-ctigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality cov.





Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 393)  
AUTHORS  
Olivier, M. and Cox, D.R.  
TITLE  
Unpublished, Olivier, M., Cox, D.R. (2000)  
JOURNAL  
Unpublished (2000)  
COMMENT

Contact: Michael Olivier, David R. Cox  
Stanford Human Genome Center  
Stanford University School of Medicine  
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA  
Tel: (650) 320-5800  
Fax: (650) 320-5801

Email: olivier@shgc.stanford.edu  
Primer A: AGAGGATTGCTTATCTCTCCCG  
Primer B: ACGGTACGATGATCAAGATGGAG  
STS size: 290  
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700

Protocol:

Template: 25 ng  
each 1 uM  
Primer: each 200 uM  
dNTPs: each 200 uM  
AmpliTaq Gold Polymerase: 0.07 units/uL  
Total Vol: 5 uL

Buffer:

MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

Finished human sequence in NCBI. STSS designed and developed at the  
Stanford Human Genome Center.

FEATURES  
source

1..393  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="13"  
/clone\_lib="Human"  
12..301

primer\_bind  
12..34

primer\_bind  
complement(279..301) 115 t

BASE COUNT

80 a 105 c 93 g 115 t

ORIGIN

Query Match 14.1%; Score 162.4; DB 11; Length 393;  
Best Local Similarity 95.9%; Pred. No. 2.2e-28;  
Matches 188; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 790 GCGTGTGCAACA-CCACCCGCGACGCCGACTTCAGGAGCTGCCGAGCCGCTTTG 848  
|||||  
Db 393 GCGTGTGCAACAACCCCGCGACGCCGACTTCAGGAGCTGCCGAGCCGCTTTG 334

QY 849 GCAGAGATCACGAGATGTTGGCACAGC-TCCACGGTAGCATGATCAAGATGGAAATTT 907  
|||||  
Db 333 GCAGAGATCACGAGATGTTGGCACAGCTTCCACGGTAGCATGATCAAGATGGAAATTT 274

QY 908 CCAGAAGCTGCAGGAATCAAGAAAGATTTGATGTCATGCAATCTTGTGTTCCCGG 967  
|||||  
Db 273 CCAGAAGCTGCAGGAATCAAGAAAGATTTGATGTCATGCAATCTTGTGTTCCCGG 214

QY 968 AAGGGAGTTTCATCCGT 983  
|||||  
Db 213 AAGGGTAAGCAGCAGT 198

RESULT 12

AX261839/c

LOCUS

DEFINITION Sequence 1490 from Patent WO0173027.

ACCESSION AX261839

VERSION AX261839.1 GI:16510806

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Meagher, M.J., Xu, J. and King, G.E.

TITLE Compositions and methods for therapy and diagnosis of colon cancer

JOURNAL Patent: WO 0173027-A 1490 04-OCT-2001;

Corixa Corporation (US)

FEATURES

Location/Qualifiers

source 1..426

/organism="Homo sapiens"

/db\_xref="taxon:9606"

BASE COUNT 89 a 109 c 109 g 119 t

ORIGIN

Query Match 13.4%; Score 154.6; DB 6; Length 426;

Best Local Similarity 61.9%; Pred. No. 1.6e-26;

Matches 263; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

QY 321 TACTTCATAGCTAAGGAAGTGTCTACCCGAGCGAACAATATCTGAAGATCTCGAAGTT 380  
|||||

Db 426 TACTTCATAGCTAAGGAAGTGTCTCGCTACAGAACGAAACATCTCAAGATTTAGAAGTT 367  
|||||

QY 381 ATCACTTTCGTGTTTCAGAGCACAGTGTAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAA 440  
|||||

Db 366 ATTACCGTGTGTTCCGCGAGCGAGTGTGAGGAGGAGCGCCATGCCTCGACTCTGATG 307  
|||||

QY 441 AGTCATATATCCGAAATTTGAACCTTTGACAAATTTTCATACATTTTCTCAAGGAA 500  
|||||

Db 306 AGCGTCTCTTCCAAACATCGATCCATCTATGATTCACAGAGGCTTCCTGCGCGAG 247  
|||||

QY 501 ATTGAGCAACGACTTGCCTGTGGGAGCGCTCAATGCCCAATACAG-ATTAC 557  
|||||

Db 246 GTGGAGCAGAGCTGGCACTCTGGGAGGCGCTCCAAAGGCCACACAAAGGAGCTAT 187  
|||||

QY 558 CAAAGAATCGCGATGTCTATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTAC 617  
|||||

Db 186 CAACGAATCGGGACATCTCTGCTCAGGAACATGCGCCAGTTAAAGGAGTTTACCAGCTAC 127  
|||||

QY 618 CTGTGGAGCACAGCGAGCCCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGCGG 677  
|||||

Db 126 TTCCAAAGACATGACGAGATCTTAACAGAACTGGAAGGCTACCAAGCGTGTAAAGAG 67  
|||||

QY 678 CTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGCTGTGTACCTACCGCTCAACACC 737  
|||||

Db 66 TTGGAGGAGCTGTACAGGAGTTTGGCTGCAGAGGCTGTCTACTTGCCTCTCAACAGC 7  
|||||

QY 738 TTCTCT 742

Db 6 TTCTCT 2

RESULT 13

AC130919/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-159A1, \*\*\* SEQUENCING IN PROGRESS

AC130919 120911 bp DNA linear HTG 15-AUG-2002

\*\*\*, 63 unordered pieces.

ACCESSION AC130919

VERSION AC130919.1 GI:22261960

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

1 (bases 1 to 120911)

## AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagm,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Gear,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,I., Haviak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kwis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuwha,E., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarunnsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reese,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Speed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wel,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished  
2 (bases 1 to 120911)  
Rat Genome Sequencing Consortium.  
Submitted (13-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: KBGN

Center clone name: CH230-159A1

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 53051 bases at least Q40

Consensus quality: 57784 bases at least Q30

Consensus quality: 61330 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently consists of 63 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 1028: contig of 1028 bp in length  
\* 1128: gap of unknown length  
\* 1129 1128: contig of 1037 bp in length  
\* 2166 2165: gap of unknown length  
\* 3620 3620: contig of 1355 bp in length  
\* 3621 3620: gap of unknown length  
\* 3721 5093: contig of 1373 bp in length  
\* 5094 5193: gap of unknown length  
\* 5194 5193: contig of 1387 bp in length  
\* 6581 6580: gap of unknown length  
\* 6681 8169: contig of 1489 bp in length  
\* 8170 8269: gap of unknown length  
\* 8270 8977: contig of 1628 bp in length  
\* 8988 9997: gap of unknown length  
\* 9998 11330: contig of 1333 bp in length  
\* 11331 11430: gap of unknown length  
\* 11431 13084: contig of 1654 bp in length  
\* 13085 13184: gap of unknown length  
\* 13185 14532: contig of 1348 bp in length  
\* 14533 14632: gap of unknown length  
\* 14633 15891: contig of 1259 bp in length  
\* 15892 15991: gap of unknown length  
\* 15993 17101: contig of 1110 bp in length  
\* 17102 17201: gap of unknown length  
\* 17202 18776: contig of 1575 bp in length  
\* 18777 18876: gap of unknown length  
\* 18877 20138: contig of 1262 bp in length  
\* 20139 20238: gap of unknown length  
\* 20239 21285: contig of 1047 bp in length  
\* 21286 21385: gap of unknown length  
\* 21386 26330: contig of 1245 bp in length  
\* 26331 27230: gap of unknown length  
\* 27231 23783: contig of 1053 bp in length  
\* 23784 23883: gap of unknown length  
\* 23884 25084: contig of 1201 bp in length  
\* 25085 25184: gap of unknown length  
\* 25185 26257: contig of 1073 bp in length  
\* 26258 26357: gap of unknown length  
\* 26358 27942: contig of 1385 bp in length  
\* 27943 28042: gap of unknown length  
\* 28043 29312: contig of 1270 bp in length  
\* 29313 29412: gap of unknown length  
\* 29413 30808: contig of 1396 bp in length  
\* 30809 30908: gap of unknown length  
\* 30909 32522: contig of 1614 bp in length  
\* 32523 32623: gap of unknown length  
\* 32623 34182: contig of 1560 bp in length  
\* 34183 34282: gap of unknown length  
\* 34283 36102: contig of 1820 bp in length  
\* 36103 36202: gap of unknown length  
\* 36203 37338: contig of 1136 bp in length  
\* 37339 37438: gap of unknown length  
\* 37439 38559: contig of 1121 bp in length  
\* 38560 38659: gap of unknown length  
\* 38660 40454: contig of 1795 bp in length  
\* 40455 40554: gap of unknown length  
\* 40555 42228: contig of 1674 bp in length  
\* 42229 42329: gap of unknown length  
\* 42329 43442: contig of 1114 bp in length  
\* 43443 43542: gap of unknown length  
\* 43543 45213: contig of 1671 bp in length  
\* 45214 45313: gap of unknown length  
\* 45314 47351: contig of 2038 bp in length  
\* 47352 47451: gap of unknown length

[illegible]

Consensus quality: 186552 bases at least Q20  
Insert size: 206000; agarose-fp  
Quality size: 187888; sum-of-contigs  
Quality coverage: 5.7 in Q20 bases; agarose-fp  
Quality coverage: 6.3 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 698: contig of 698 bp in length  
699 798: gap of 100 bp  
799 1437: contig of 639 bp in length  
1438 1537: gap of 100 bp  
1538 2255: contig of 718 bp in length  
2256 2355: gap of 100 bp  
2356 3033: contig of 678 bp in length  
3034 3133: gap of 100 bp  
3134 3783: contig of 650 bp in length  
3784 3883: gap of 100 bp  
3884 4559: contig of 676 bp in length  
4560 4659: gap of 100 bp  
4660 5309: contig of 650 bp in length  
5310 5409: gap of 100 bp  
5410 6053: contig of 644 bp in length  
6054 6153: gap of 100 bp  
6154 7009: contig of 856 bp in length  
7010 7109: gap of 100 bp  
7110 7747: contig of 638 bp in length  
7748 7847: gap of 100 bp  
7848 8125: contig of 278 bp in length  
8126 8225: gap of 100 bp  
8226 8956: contig of 731 bp in length  
8957 9056: gap of 100 bp  
9057 10142: contig of 1086 bp in length  
10143 10242: gap of 100 bp in length  
10243 11404: contig of 1162 bp in length  
11405 11504: gap of 100 bp  
11505 12593: contig of 1089 bp in length  
12594 12693: gap of 100 bp  
12694 14681: contig of 1988 bp in length  
14682 14781: gap of 100 bp in length  
14782 17534: contig of 2753 bp in length  
17535 17634: gap of 100 bp  
17635 21928: contig of 4294 bp in length  
21929 22028: gap of 100 bp  
22029 29432: contig of 7404 bp in length  
29433 29532: gap of 100 bp in length  
29533 35637: contig of 6105 bp in length  
35638 35737: gap of 100 bp  
35738 84581: contig of 48844 bp in length  
84582 84681: gap of 100 bp  
84682 93581: contig of 8900 bp in length  
93582 93681: gap of 100 bp  
93682 101846: contig of 8165 bp in length  
101847 101946: gap of 100 bp  
101947 111400: contig of 9434 bp in length  
111401 111500: gap of 100 bp  
111501 131247: contig of 19747 bp in length  
131248 131347: gap of 100 bp  
131348 151032: contig of 19685 bp in length  
151033 151132: gap of 100 bp  
151133 184681: contig of 33549 bp in length  
184682 184781: gap of 100 bp  
184782 190588: contig of 5807 bp in length.

FEATURES  
Source

1. .190588  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

/clone="RP23-321M11"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
1. .698  
/note="assembly\_fragment"  
799. .1437  
/note="assembly\_fragment"  
1538. .2255  
/note="assembly\_fragment"  
2356. .3033  
/note="assembly\_fragment"  
3134. .3783  
/note="assembly\_fragment"  
3884. .4559  
/note="assembly\_fragment"  
4660. .5309  
/note="assembly\_fragment"  
5410. .6053  
/note="assembly\_fragment"  
6154. .7009  
/note="assembly\_fragment"  
7110. .7747  
/note="assembly\_fragment"  
7848. .8125  
/note="assembly\_fragment"  
8226. .8956  
/note="assembly\_fragment"  
9057. .10142  
/note="assembly\_fragment"  
10243. .11404  
/note="assembly\_fragment"  
11505. .12593  
/note="assembly\_fragment"  
12694. .14681  
/note="assembly\_fragment"  
14782. .17534  
/note="assembly\_fragment"  
17635. .21928  
/note="assembly\_fragment"  
22029. .29432  
/note="assembly\_fragment"  
29533. .35637  
/note="assembly\_fragment"  
35738. .84581  
/note="assembly\_fragment"  
84682. .93581  
/note="assembly\_fragment"  
93682. .101846

Query Match 12.9%; Score 148.6; DB 2; Length 190588;  
Best Local Similarity 83.3%; Pred. No. 6e-25;  
Matches 169; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 97 CGCAGCCACGACAGGCTCCCTGACTGGCAGTCTCTACCTTTCCGAGCTGCTGTGNACT 156  
Db 187607 CTATCCCCAACACAGGCTCCCTGACTGGTAGCCCTCACCTTTTCAGAGCTGCTCATCACT 187548  
QY 157 CGCAGGGGGAGTGCGCCCTGCCAACGTGACCTTGCTCCCAACCTGAGCCCGCACCA 216  
Db 187547 CACAGGGAGGAGCGCGCCCGCAGCAGCAACGTGACCTGTCTCCCAACCTGAGTCCTGACACA 187488  
QY 217 AGCAGGCTCTCCCTTGTATCAGCCCGCTGCTGAATGACCAAGGCTGCCCGGACGAGC 276  
Db 187487 AGCAGGCTCTCCCTGTATCAGCCCGCTGCTGAACGACCAAGCGTGCCCAAGGACGAGC 187428  
QY 277 ATCAGATGAGGCGCGGGAAG 299  
Db 187427 ACGAAGAAGAGCGCGCGGAAG 187405

RESULT 15  
AC122885/c  
LOCUS  
DEFINITION Mus musculus chromosome UNK clone RP23-90H14, WORKING DRAFT  
AC122885 221502 bp DNA linear HTG 23-JUN-2002  
Mus musculus chromosome UNK clone RP23-90H14, WORKING DRAFT

SEQUENCE, 8 unordered pieces.  
AC122885  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 221502)  
McPherson, J.D. and Waterston, R.H.  
The sequence of Mus musculus clone  
Unpublished  
2 (bases 1 to 221502)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 221502)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
On Jun 23, 2002 this sequence version replaced gi:21218511.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@wustl.wustl.edu  
----- Project Information -----  
Center project name: M\_BA0090H14  
----- Summary Statistics -----

Sequencing vector: W13; 0%  
Sequencing vector: plasmid; 100%  
Chemistry: Dye-primer ET; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 216976 bases at least Q40  
Consensus quality: 217754 bases at least Q30  
Consensus quality: 218341 bases at least Q20  
Insert size: 190000; agarose-fp  
Insert size: 221868; sum-of-contigs  
Quality coverage: 15.60 in Q20 bases; agarose-fp  
Quality coverage: 12.41 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

1 8441: contig of 8441 bp in length  
\* 8442 8541: gap of unknown length  
\* 8542 27277: contig of 18736 bp in length  
\* 27278 27377: gap of unknown length  
\* 27378 42339: contig of 14962 bp in length  
\* 42340 42439: gap of unknown length  
\* 42440 71538: contig of 29099 bp in length  
\* 71539 99587: gap of unknown length  
\* 99588 99687: contig of 27949 bp in length  
\* 99688 134048: gap of unknown length  
\* 134049 134148: gap of unknown length  
\* 134149 221349: contig of 87201 bp in length  
\* 221350 221449: gap of unknown length  
\* 221450 221502: contig of 53 bp in length.

Location/Qualifiers  
1..221502  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

FEATURES  
Source

/chromosome="UNK"  
/clone="RP23-90H14"  
1..8441  
/note="assembly\_name:Contig59"  
8542..27277  
/note="assembly\_name:Contig60"  
27378..42339  
/note="assembly\_name:Contig61"  
42440..71538  
/note="assembly\_name:Contig62"  
71639..99587  
/note="assembly\_name:Contig63"  
99688..134048  
/note="assembly\_name:Contig64"  
134149..221349  
/note="assembly\_name:Contig65"  
221450..221502  
/note="assembly\_name:Contig19"  
BASE COUNT 61405 a 54522 c 50450 g 54373 t 752 others  
ORIGIN  
Query Match 12.9%; Score 148.6; DB 2; Length 221502;  
Best Local Similarity 83.3%; Pred. No. 6.1e-25;  
Matches 169; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
QY 97 CGCAGCCAAAGCAGCAGGCTCCCTGACTGGCAGTCCTCACCTTCCGAGCTGTCTGTGAAC 156  
Db 113505 CTATCCCCAAAGAGGCTCCCTGACTGGTAGCCCTCACCTTTTCAGAGCTGTCCATCAAC 113446  
QY 157 CGCAGGGGGAGTGGCCCTGCGCAAGTGACCTTGTCTCCCAACCTGAGCCCGACACCA 216  
Db 113445 CACAGGAGGAGCGCCGCCAGCAAGTGACCTGTCTCCCAACCTGAGTCTGTGACAACA 113386  
QY 217 AGCAGGCTCTCCCTTGATCAGCCCTGCTGANTGACAGGCTGCCCGGAGCGAGC 276  
Db 113385 AGCAGGCTCTCCCTTGATCAGCCCTGCTGANTGACAGGCTGCCCGGAGCGAGC 113326  
QY 277 ATGAGGATGAGGGCGGAGGAAG 299  
Db 113325 ACGAAGAGAGGGCGGCGGAAG 113303  
Search completed: December 3, 2002, 21:05:54  
Job time : 3209.8 secs

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 14:48:04 : Search time 181.787 Seconds  
(without alignments)  
14246.337 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_1352\_2501

Perfect score: 1150

Sequence: 1 cgggtacaagcaggcgac.....attgaggagagcgaagacga 1150

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1150	100.0	3442	20	AAAX79183 Human chondrocyte-
2	1150	100.0	4687	21	AAAO8582 Human cytoskeleton
3	1146.8	99.7	3187	23	AAAS64760 DNA encoding novel
4	762	66.3	3094	21	AAC98992 Human pancreatic c
c 5	154.6	13.4	426	23	AAAS58814 cDNA #1490 encodin
c 6	112	9.7	1824	23	AAAS64761 DNA encoding novel
7	106	9.2	1718	22	AAAS27016 cDNA encoding nove
8	105.8	9.2	1747	23	ABLO3867 Drosophila melanog
c 9	105.8	9.2	4150	23	ABLO3866 Drosophila melanog

10	105.2	9.1	716	22	AAS27440 cDNA encoding nove
11	98	8.5	717	24	ABL90828 Human polynucleoti
12	90.2	7.8	544	22	AAH98023 Murine 7-transmemb
13	83.6	7.3	665	22	AAH07867 Human cDNA clone (
14	83.6	7.3	1296	24	ABK49898 Human cDNA encodin
15	83.6	7.3	1821	22	AAH15014 Human cDNA sequenc
16	83.6	7.3	3610	24	ABN59933 Novel human coding
c 17	72.8	6.3	462	24	ABK53702 Human eosinophil-m
18	64.4	5.6	2686	22	AAAS27013 cDNA encoding nove
19	64.4	5.6	3321	22	AAAS44718 Human full-length
20	64	5.6	1091	22	AAAS27281 cDNA encoding nove
21	60.8	5.3	1860	22	AAAF9965 Human colon carcin
22	60.8	5.3	5923	22	AAAS02052 DNA encoding molec
c 23	60.8	5.3	5923	22	AAAS02052 DNA encoding molec
24	60	5.2	60	24	ABN41498 Human spliced tran
25	56.2	4.9	365	14	AAQ39923 Expressed Sequence
26	56.2	4.9	365	14	AAQ59335 Human brain Expres
27	53.2	4.6	417	22	ABA08384 Human collybistin
c 28	46.4	4.0	454	23	ABV54945 Human prostate exp
29	46.4	4.0	541	24	ABN64838 Human cancer relat
30	44.4	3.9	342	21	AAF08796 Fusarium venenatum
31	44	3.8	1609	23	ABL25953 Drosophila melanog
32	44	3.8	3381	23	ABL25950 Drosophila melanog
33	44	3.8	3609	23	ABL25952 Drosophila melanog
34	43.4	3.8	3786	23	AAAS4228 Pseudomonas aerugi
c 35	42.4	3.7	1730	22	AAAS27017 cDNA encoding nove
36	42.4	3.7	1730	22	AAAS27441 cDNA encoding nove
37	42.4	3.7	1730	22	AAI62819 Human cDNA SEQ ID
c 38	42.4	3.7	1778	22	AAI60338 Human polynucleoti
39	42.4	3.7	1816	22	AAI58552 Human polynucleoti
40	42	3.7	3306	23	ABL12855 Drosophila melanog
41	41.2	3.6	2159	15	AAQ67223 Mouse p53Nuc. Mus
42	40.8	3.5	625	22	AAH07308 Human cDNA clone (
43	40.8	3.5	1444	22	AAH41863 Human chondroadher
44	40.8	3.5	1696	22	AAI93904 Human stomach canc
45	40.8	3.5	1696	22	AAH16459 Human cDNA sequenc

## ALIGNMENTS

RESULT 1  
AAAX79183  
ID AAAX79183 standard; DNA; 3442 BP.

XX

AC AAAX79183;

XX 17-AUG-1999 (first entry)

XX Human chondrocyte-derived gene CDEP.

DE Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;

XX Dbl homology domain; pleckstrin homology domain; rheumatoid arthritis;

KW drug; ss.

XX Homo sapiens.

XX OS

XX PN WO9928458-A1.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98WO-JP05348.

XX PR 27-NOV-1997; 97JP-0342060.

XX (CHUS ) CHUGAI SEIYAKU KK.

XX PI Kato Y, Kawamoto T, Koyano Y;

XX WPI: 1999-371117/31.

XX P-PSDB; AAY07482.

XX PT Protein CDEP expressed in differentiated chondrocytes, and gene





CC reproductive and muscle disorders. Pharmaceutical compositions  
CC containing CYSP-2 to CYSP-16 can be used to treat or prevent disorders  
CC associated with decreased expression or activity of CYSP (claimed), for  
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,  
CC cancers, autoimmune/anti-inflammatory disorders such as allergies, anemia,  
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,  
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma  
CC and trauma. CYSP antagonists can be used to treat or prevent a  
CC disorder associated with increased expression or activity of CYSP  
CC (claimed).

XX  
SQ Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;

Query Match 100.0%; Score 1150; DB 21; Length 4687;  
Best Local Similarity 100.0%; Pred. No. 8.2e-307;  
Matches 1150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGTAACAGCAGCGGCGAGCGGCTCGGCGCCACGAGGAGGAGGAGGTGG 60  
DB 1393 CGGGTAACAGCAGCGGCGAGCGGCTCGGCGCCACGAGGAGGAGGAGGTGG 1452

QY 61 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCCGAGGAGGAGGAGGTGG 120  
DB 1453 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCCGAGGAGGAGGAGGTGG 1512

QY 121 CTGGAGTCTCACCCTTCCGAGCTCTCTGTGAATCTCGAGGGGAGTGGCCCTGCCA 180  
DB 1513 CTGGAGTCTCACCCTTCCGAGCTCTCTGTGAATCTCGAGGGGAGTGGCCCTGCCA 1572

QY 181 ACGTGACCTTGCTCCCAACTGAGCCCGGACACCAAGCAGGCGCTCTCCCTTGATCAGCC 240  
DB 1573 ACGTGACCTTGCTCCCAACTGAGCCCGGACACCAAGCAGGCGCTCTCCCTTGATCAGCC 1632

QY 241 CGCTGCTGAATGACAGGCTCGCCCGGACGAGATGAGGATGAGGCGCGGAGGAAGA 300  
DB 1633 CGCTGCTGAATGACAGGCTCGCCCGGACGAGATGAGGATGAGGCGCGGAGGAAGA 1692

QY 301 GATTCCTCACTGATTAAGCGTACTTCATAGCTAAGGAAGTGTCTACCCAGCGGCAACAT 360  
DB 1693 GATTCCTCACTGATTAAGCGTACTTCATAGCTAAGGAAGTGTCTACCCAGCGGCAACAT 1752

QY 361 AYTCTGAAGATCTGAGTATATCTCTGTTGTTTTCAGAGCAGTGTGCAAGAGGAGG 420  
DB 1753 ATCTGAAGATCTGAGTATATCTCTGTTGTTTTCAGAGCAGTGTGCAAGAGGAGG 1812

QY 421 CCATGCCGGAAGCAGTGAAGTCTCATATCCCGAATTTTGAACCTTTGCACAAATTC 480  
DB 1813 CCATGCCGGAAGCAGTGAAGTCTCATATCCCGAATTTTGAACCTTTGCACAAATTC 1872

QY 481 ATACTAATTTTCTAAGGAAATTTGAGCAACGACTTGCCCTGTGGGAGGCGCTCAAAATG 540  
DB 1873 ATACTAATTTTCTAAGGAAATTTGAGCAACGACTTGCCCTGTGGGAGGCGCTCAAAATG 1932

QY 541 CCCAATCAGATTTACCAAGAAATCGGGGATGTGATGTGAGAACATTCAGGGCATGA 600  
DB 1933 CCCAATCAGATTTACCAAGAAATCGGGGATGTGATGTGAGAACATTCAGGGCATGA 1992

QY 601 AGCACCTGGCGCTACCTGTGSAAGCAGCAGCGGCTTGAGGCGCTTGAGAAATGGA 660  
DB 1993 AGCACCTGGCGCTACCTGTGSAAGCAGCAGCGGCTTGAGGCGCTTGAGAAATGGA 2052

QY 661 TCAAGAGCTCCGCGGCTGGAGAACTTCTGAGAGACTTTGAGCTGCAAGAGGTGTGTT 720  
DB 2053 TCAAGAGCTCCGCGGCTGGAGAACTTCTGAGAGACTTTGAGCTGCAAGAGGTGTGTT 2112

QY 721 ACCTACCGCTCAACAGCTTCTCTCGGCGCACTGACCGGCTCATGCACTACAAGCAGG 780  
DB 2113 ACCTACCGCTCAACAGCTTCTCTCGGCGCACTGACCGGCTCATGCACTACAAGCAGG 2172

QY 781 TCTGAGCGGCTGTGCAACACCAACCCCGCCGAGCCAGCTTCAGGAGTCCCGAG 840  
DB 2173 TCTGAGCGGCTGTGCAACACCAACCCCGCCGAGCCAGCTTCAGGAGTCCCGAG 2232

QY 841 CCGCTTTGCGAGATACGAGAGTGGTGGCACAGCTCCACGATGATCAAGATGG 900  
DB 2233 CCGCTTTGCGAGATACGAGAGTGGTGGCACAGCTCCACGATGATCAAGATGG 2292

QY 901 AGAATTTCCAGAGCTGCACGAACCTCAAGAAAGATTGATTGGCATTGACAAATCTTGTGG 960  
DB 2293 AGAATTTCCAGAGCTGCACGAACCTCAAGAAAGATTGATTGGCATTGACAAATCTTGTGG 2352

QY 961 TTCCGGGAAGGAGTTTCAATCCCTCTGGGCGAGCTCAGCAAGCTCTCGGGGAAAGGGCTCC 1020  
DB 2353 TTCCGGGAAGGAGTTTCAATCCCTCTGGGCGAGCTCAGCAAGCTCTCGGGGAAAGGGCTCC 2412

QY 1021 AGCAGCGCATGTTCTTCTGTTTCAACGAGCTCTGCTATACAGAGCCCGGGGCTGACGG 1080  
DB 2413 AGCAGCGCATGTTCTTCTGTTTCAACGAGCTCTGCTATACAGAGCCCGGGGCTGACGG 2472

QY 1081 CCTCAATCAGTTTAAAGTCCACGGGCGAGCTCCCGCTCTATGGCATGAGGATGAGGAGA 1140  
DB 2473 CCTCAATCAGTTTAAAGTCCACGGGCGAGCTCCCGCTCTATGGCATGAGGATGAGGAGA 2532

QY 1141 GCGAAGACGA 1150  
DB 2533 GCGAAGACGA 2542

RESULT 3  
AAS64760  
ID AAS64760 standard; cDNA; 3187 BP.  
XX AAS64760;  
XX AC  
XX 13-FEB-2002 (first entry)  
XX DNA encoding novel human diagnostic protein #564.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG00573.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX Claim 1; SEQ ID No 564; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 3187 BP; 810 A; 874 C; 876 G; 627 T; 0 other:  
Query Match 99.7%; Score 1146.8; DB 23; Length 3187;  
Best Local Similarity 99.8%; Pred. No. 5.2e-306;  
Matches 1148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGGTAACAGGAGCGGAGCGCGCTCGCGCCGCCACCGAGGAAGAGGAGGTCG 60  
DB 1353 CGGGTAACAGGAGCGGAGCGCGCTCGGAGCCCGAGGAGGAGGTCG 1412  
QY 61 TTAAGATAGGACCCAGCAGAGATAACCTCAGCCCGCCGAGCAAGCAGAGCTCCCTGA 120  
DB 1413 TTAAGATAGGACCCAGCAGAGATAACCTCAGCCCGCCGAGCAAGCAGAGCTCCCTGA 1472  
QY 121 CTGGCAGTCTCAGCTTCCGAGCTGCTGTGAACCTCGCAGGGGGAGTGGCCCTGGCA 180  
DB 1473 CTGGCAGTCTCAGCTTCCGAGCTGCTGTGAACCTCGCAGGGGGAGTGGCCCTGGCA 1532  
QY 181 AGTGACCTTGCTCCCAACCTGAGCCCGCAGCCACCAAGCAGGCGCTCTCCCTTGATCAGCC 240  
DB 1533 AGTGACCTTGCTCCCAACCTGAGCCCGCAGCCACCAAGCAGGCGCTCTCCCTTGATCAGCC 1592  
QY 241 CGCTGCTGAATGACAGCGCTCGCCCGGAGCGAGATGAGGATGAGGCGCGGAGGAAGA 300  
DB 1593 CGCTGCTGAATGACAGCGCTCGCCCGGAGCGAGATGAGGATGAGGCGCGGAGGAAGA 1652  
QY 301 GATTCCTCACTGATAAGCGTACTCATAGCTAAGGAGTGTCTACCCAGCGAGCAACAT 360  
DB 1653 GATTCCTCACTGATAAGCGTACTCATAGCTAAGGAGTGTCTACCCAGCGAGCAACAT 1712  
QY 361 ATCTGAAGGATCTGAAAGTTATCACTTCGTGTTTCAGAGCAGTGTAGCAAGAGGAGCG 420  
DB 1713 ATCTGAAGGATCTGAAAGTTATCACTTCGTGTTTCAGAGCAGTGTAGCAAGAGGAGCG 1772  
QY 421 CCATGCGGAGCACTGAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCAACAATTTTC 480  
DB 1773 CCATGCGGAGCACTGAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCAACAATTTTC 1832  
QY 481 ATACTAATTTTCTCAAGGAAATTTGACACGACTTGCCTGTGGGAAGCGCTCAAAATG 540  
DB 1833 ATACTAATTTTCTCAAGGAAATTTGACACGACTTGCCTGTGGGAAGCGCTCAAAATG 1892  
QY 541 CCCAATCAGAGATTACCAAGAATCGGCGATGTCATGCTGAAGAACATTTTCAGGCGATGA 600  
DB 1893 CCCAATCAGAGATTACCAAGAATCGGCGATGTCATGCTGAAGAACATTTTCAGGCGATGA 1952  
QY 601 AGCACTGCGCGCTCACCTGTGGAAGCAGCAGCGAGCGCTTGGAGGCGCTTGGAGAATGGA 660  
DB 1953 AGCACTGCGCGCTCACCTGTGGAAGCAGCAGCGAGCGCTTGGAGGCGCTTGGAGAATGGA 2012  
QY 661 TCAAGAGCTCCCGCGGCTGGAGAACTTCTGAGAGACTTTGAGCTGAGAGAGTGTGTT 720  
DB 2013 TCAAGAGCTCCCGCGGCTGGAGAACTTCTGAGAGACTTTGAGCTGAGAGAGTGTGTT 2072  
QY 721 ACCTACCGCTCAACACTTCCCTCGCGCCACTTGACCGCGCTCATGCACTACAAGCAGG 780  
DB 2073 ACCTACCGCTCAACACTTCCCTCGCGCCACTTGACCGCGCTCATGCACTACAAGCAGG 2132  
QY 781 TCGTGAGCGGCTGTGCAAAACACACCCCGGAGCCAGCCGACTTCAGGAGTCCCGAG 840  
DB 2133 TCGTGAGCGGCTGTGCAAAACACACCCCGGAGCCAGCCGACTTCAGGAGTCCCGAG 2192

QY 841 CGCTTTGGCAGAGATCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGG 900  
DB 2193 CGCTTTGGCAGAGATCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGG 2252  
QY 901 AGAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGG 960  
DB 2253 AGAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGG 2312  
QY 961 TTCGGGAAGGAGTTTCATCGTCTGGCAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCC 1020  
DB 2313 TTCGGGAAGGAGTTTCATCGTCTGGCAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCC 2372  
QY 1021 AGCAGCGATGTTCTTCTTCAACGAGCTCTCTATACAGAGCCGGGGCTGACGG 1080  
DB 2373 AGCAGCGATGTTCTTCTTCAACGAGCTCTCTATACAGAGCCGGGGCTGACGG 2432  
QY 1081 COTCCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAGA 1140  
DB 2433 COTCCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAGA 2492  
QY 1141 GCGAAGACGA 1150  
DB 2493 GCGAAGACGA 2502

RESULT 4  
AAC98992  
ID AAC98992 standard; cDNA; 3094 BP.  
XX AC AAC98992;  
XX AC AAC98992;  
DT 09-MAR-2001 (first entry)  
XX  
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.  
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antinflammatory; cardiant; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW pulmonary; cardiovascular; renal; proliferative; ss.  
XX Homo sapiens.  
XX WO200055320-A1.  
XX 21-SEP-2000.  
XX 08-MAR-2000; 2000WO-US05989.  
XX 12-MAR-1999; 99US-0124270.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX WPI; 2000-579444/54.  
XX P-PSDB; AAB54227.  
PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
PS Claim 1; Page 664-665; 1379pp; English.  
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for

preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer. Agonists and antagonists to the antigens can be screened for. The pancreatic cancer antigen polynucleotides can be used to design nucleic acid hybridization probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of forensic and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both *in vivo* and *in vitro* diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent sequences used in the exemplification of the present invention.

Sequence 3094 BP; 736 A; 849 C; 755 G; 749 T; 5 other;

Query Match 66.3%; Score 762; DB 21; Length 3094;  
Best Local Similarity 100.0%; Pred. NO. 8.5e-200;  
Matches 762; Conservative 0; Mismatches 0; Indels 0

389	GTGGTTTTCAGAGCACAGTGCAGCAAGAGGAGCGCATCCCGGAGAGCACTGAAAAGTCTCAT	448
62	GTGGTTTTCAGAGCACAGTGCAGCAAGAGGAGCGCATCCCGGAGAGCACTGAAAAGTCTCAT	121
449	ATTCCCGAATTTTGAACCTTTGCACAATTTTCACTAATTTTCTCAAGGAAATTCAGCA	508
122	ATTCCCGAATTTTGAACCTTTGCACAATTTTCACTAATTTTCTCAAGGAAATTCAGCA	181
509	ACGACTTGCCTGTGGGAAGCGCCTCAATGCCCCAAATCAGAGATTACCAAGAATTCGG	568
182	ACGACTTGCCTGTGGGAAGCGCCTCAATGCCCCAAATCAGAGATTACCAAGAATTCGG	241
569	CGATGTCATGCTGAAGAACATTTACAGGCATCAAGCACTCGCGGCTCACTGTGGAAGCA	628
242	CGATGTCATGCTGAAGAACATTTACAGGCATCAAGCACTCGCGGCTCACTGTGGAAGCA	301
629	CAGCGAGGCCCTTGGAGGCCCTTGGAGAATGGAATCAAGAGTCCCGGCGCTGGAGAACTT	688
302	CAGCGAGGCCCTTGGAGGCCCTTGGAGAATGGAATCAAGAGTCCCGGCGCTGGAGAACTT	361
689	CTGCAGAGACTTTGAGCTGCAGAGAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGG	748
362	CTGCAGAGACTTTGAGCTGCAGAGAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGG	421
749	GCCACTGCACCGGCTCATGCACCTACAAGCAGGTCTTGGAGCGCTGTGCAAAACACCAACC	808
422	GCCACTGCACCGGCTCATGCACCTACAAGCAGGTCTTGGAGCGCTGTGCAAAACACCAACC	481
809	CGCGAGCCACGCGACTTCAGGGACTCCGAGCGGCTTTGGCAGAGATCACGGAGATGT	868
482	CGCGAGCCACGCGACTTCAGGGACTCCGAGCGGCTTTGGCAGAGATCACGGAGATGT	541
869	GGCAGAGCTCACAGTACGATGATCAAGATGGAGAAATTTCCAGAAGCTGCACGAATCAA	928
542	GGCAGAGCTCACAGTACGATGATCAAGATGGAGAAATTTCCAGAAGCTGCACGAATCAA	601
929	GAAGATTGTGATTGGCAATTGACAATCTTGTGGTTCCGGGAAGGAGTTCATCCGCTCGGG	988
602	GAAGATTGTGATTGGCAATTGACAATCTTGTGGTTCCGGGAAGGAGTTCATCCGCTCGGG	661
989	CAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTGTCAACGA	1048
662	CAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTGTCAACGA	721
1049	CGTCTGCTATACACGAGCGGGGCTGACGGCTCCAACTAGTTTAAAGTCCACGGGCA	1108
722	CGTCTGCTATACACGAGCGGGGCTGACGGCTCCAACTAGTTTAAAGTCCACGGGCA	781
1109	GCTCCCGCTCTATTGGCATGACGATTGAGGAGAGCGAAGCGA	1150

Db 782 GCTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 823

## RESULT 5

AA58814/C	
ID	AA58814 standard; cDNA; 426 BP.
XX	
AC	AA58814;
XX	
DT	13-FEB-2002 (first entry)
XX	
XX	
DE	cDNA #1490 encoding portion of a human colon tumour protein.
XX	
KW	Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200173027-A2.
XX	
PD	04-OCT-2001.
XX	
PF	22-MAR-2001; 2001WO-US09246.
XX	
PR	24-MAR-2000; 2000US-191597P.
PR	04-MAY-2000; 2000US-202024P.
PR	05-MAY-2000; 2000US-202189P.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Meagher MJ, Xu J, King GE;
XX	
DR	WPI; 2001-611627/70.
XX	
PT	New colon tumour proteins and related nucleic acid, useful for
PT	treatment, prevention, diagnosis and monitoring of cancer -
XX	
PS	Claim 4; Page 290; 299pp; English.

Th present invention relates to the isolation of novel cDNA sequences encoding for at least an immunogenic portion of human colon tumour proteins. The sequences of the invention are useful in pharmaceutical compositions and vaccines for the prevention and treatment of cancers such as colon cancer. They are also useful for the diagnosis and monitoring of such cancers. Antibodies to the colon tumour proteins and antigen presenting cells that express polynucleotides encoding colon tumour proteins can be used to inhibit the development of cancers. T-cells that react specifically with colon tumour proteins are useful for removing tumour cells from samples (e.g. blood) and for cancer treatment. The polynucleotides sequences are also useful in gene therapy. AAS57325-AAS5880 represent the cDNA sequences of the invention that encode for portions of human colon tumour proteins.

Sequence 426 BP; 89 A; 109 C; 109 G; 119 T; 0 other;  
SQ

Query Match	13.4%	Score 154.6;	DB 23;	Length 426;
Best Local Similarity	61.9%	Pred. No. 1.5e-32;		
Matches 263;	Conservative	0;	Mismatches 159;	Indels 13;
Gaps	1;			

[illegible]



PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254097.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-465460/50.  
DR P-PSDB; AAU17099.  
XX  
PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
PT prognosing disorders related to the proteins, including cancers, immune  
XX disorders and neuronal disorders -  
XX  
PS Claim 1; SEQ ID No 51; 880pp; English.  
XX  
CC The invention relates to novel isolated polypeptides (I), and  
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
CC diagnosing, preventing and treating diseases including immune system  
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
CC transplant rejections and graft versus host disease, infectious diseases  
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
CC disorders, primary haematopoietic disorders, hyperproliferative  
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal  
CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
CC Addison's disease), reproductive system disorders, gastrointestinal  
CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
CC as stimulators of B-cell responsiveness to pathogens, activators of  
CC T-cells, to induce higher affinity antibodies, and as a means to induce  
CC tumour proliferation in pathologies e.g. acquired immune deficiency







PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0234999.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254097.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465460/50.  
DR P-PSDB; AAU17523.  
XX  
PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
PT prognosing disorders related to the proteins, including cancers, immune  
PT disorders and neuronal disorders -  
XX  
XX Claim 1; SEQ ID No 475; 880pp; English.  
PS  
XX The invention relates to novel isolated polypeptides (I), and  
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
CC diagnosing, preventing and treating diseases including immune system  
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
CC transplant rejections and graft versus host disease, infectious diseases

(e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and other blood-related disorders (sickle cell anaemia), myeloproliferative disorders, primary haematopoietic disorders, hyperproliferative disorders (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal abnormalities (down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmia), respiratory disorders, dermatological disorders (e.g. psoriasis), epithelial cell proliferation, endocrine disorders (e.g. Addison's disease), reproductive system disorders, gastrointestinal disorder (inflammatory disorders), liver disorders (cirrhosis), as stimulators of B-cell responsiveness to pathogens, activators of T-cells, to induce higher affinity antibodies, and as a means to induce tumour proliferation in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction pathway protein coding sequences and PCR primers of the invention.

Query Match 9.1%; Score 105.2; DB 22; Length 716;  
Best Local Similarity 64.0%; Pred. No. 8.5e-19;  
Matches 155; Conservative 2; Mismatches 85; Indels 0; Gaps 0;  
QY 909 CAGAGCTGCACCAAGATTTCATTGGCATTGACAACTTGTGCTCCGGGA 968  
DB 3 CAGAGCTGCACCAAGATTTCATTGGCATTGACAACTTGTGCTCCGGGA 968  
QY 969 AGGAGTTCATCGCTCTGGGAGGCTCAGCAAGCTCTCGGGGAGGGCTCCAGCAGCGC 1028  
DB 63 AGGAGTTCATCGCTCTGGGAGGCTCAGCAAGCTCACCAGAGGGCTCCAGCAGAGG 122  
QY 1029 ATGTTCTTCCTGTTACGACGCTCTGCTATACAGCGCGGGGCTGACGGCTCCCAAT 1088  
DB 123 ATGTTCTTCCTGTTACGACGCTCTGCTATACAGCGCGGGGCTGACGGCTCCCAAT 1088  
QY 1089 CAGTTTAAAGTCCAGCGGAGCTCCCGCTCTATGGCATGACGATTGAGGACCGAAGAC 1148  
DB 183 CACTTCGGATCCGGGGCCCTCTCCCTCCCAAGGCATGCTGGTGAAGAAGTGATAAC 242  
QY 1149 GA 1150  
DB 243 GA 244

RESULT 11  
ABL90828  
ID ABL90828 standard; cDNA; 717 BP.  
AC ABL90828;  
XX  
DT 24-MAY-2002 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 1390.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; gene; ss.  
OS Homo sapiens.  
XX  
PN WO200190304-A2.  
XX  
XX 29-NOV-2001.  
XX  
PF 18-MAY-2001; 2001WO-US16450.  
XX  
PR 19-MAY-2000; 2000US-205515P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Birse CE, Rosen CA;  
XX

DR WPI: 2002-122018/16.  
DR P-PSDB; ABB90419.  
XX  
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders.  
PT  
XX  
PS Claim 4; SEQ ID NO 1390; 2081pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABL9449-ABL90853) and proteins (ABB9040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 717 BP; 170 A; 196 C; 203 G; 144 T; 4 other;  
Query Match 8.5%; Score 98; DB 24; Length 717;  
Best Local Similarity 65.6%; Pred. No. 8.2e-17;  
Matches 143; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
QY 933 GATTTGATTGGCAATTCATCTGTGGTTCGGGAAGGAGTTTCATCGTCCGGCAGC 992  
DB 28 GACCTGGTGGCATAGAGAACCCTCATCTGCTCGCAGGAGGTTTCATCGTCCGGCAGC 87  
QY 993 CTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGAGTC 1052  
DB 88 CTTCAAGCTCACCAGGAAGGGCTGCGCAGCAGAGGATGTTTTTCTGTTCTCAGATATG 147  
QY 1053 CTGCTATACAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCACGGCAGCTC 1112  
DB 148 TTGCTGTACAGCAAGAGAGTTGCGAGGACCGCCACTTCGGATCCGGGCGCTCTT 207  
QY 1113 CCGCTCTATGGCATGACGATTGAGGAGCGAAGACGA 1150  
DB 208 CCCCTCAAGCATGCTGGTGAAGAAGTGATAACGA 245  
RESULT 12  
AAH98023  
ID AAH98023 standard; DNA; 544 BP.  
XX  
AC AAH98023;  
XX  
DT 10-OCT-2001 (first entry)  
XX  
DE Murine 7-transmembrane G-protein coupled receptor coding sequence #267.  
XX  
KW Murine; stromal stem cell; signalling; vaccine; 7TM-GPCR;  
KW 7-transmembrane G-protein coupled protein receptor; ds.  
XX  
XX Mus sp.  
XX OS  
XX PN WO200160999-A1.  
XX  
XX 23-AUG-2001.  
PD  
XX 14-FEB-2001; 2001WO-US04700.  
XX  
XX



thrombolic disease, polycystic ovarian syndrome, reduced fertility, complications of pregnancy, menstrual irregularities, hirsutism, stress incontinence and depression. The coding region of FGDL-like protein polynucleotide is useful for generating antisense oligonucleotides or ribozymes. These antisense oligonucleotides are useful for modulating FGDL-like protein gene expression. The FGDL-like protein is useful for generating antibodies against FGDL-like protein amino acid sequences and for use in various assay systems. The present sequence is the cDNA encoding the human FGDL-like protein.

Sequence 1296 BP; 415 A; 273 C; 281 G; 327 T; 0 other;

Query Match 7.3%; Score 83.6; DB 24; Length 1296; Best Local Similarity 51.3%; Pred. No. 1e-12; Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

Qy 719 TTACTACCGCTCAACACCTTCCTCCCTGGCGCCACTCCACGGCTCATGCACACTACAAGCA 778  
 Db 18 TAATCTGGCCCTCAAGCAGCTACCTGCTCAAGCCGGTTTCAGAGGATCCCCAGTACAGGCT 77  
 Qy 779 GGTCTCTGGAGCGCTGTGCACAAACACCCAGCCGCGGAGCTTCAGGGACTGCCCG 838  
 Db 78 GTTCTGCACAGATTATTTGAAGATCTCATAGAGATGCTGGAGATTACAGACACTCA 137  
 Qy 839 AGCCGCTTTGGCAGAGATCAGGAGATGGTGGCAGACTCCAGGGTACGATGATCAAGAT 898  
 Db 138 AGATGCCCTTCTGTTTATAGAGGTAGCAACACCGCCCAATGACACCATGAAGCAAGG 197  
 Qy 899 GGAGAAATTTCCAGAGCTGCACGAACTCAAGAAGATTGATTTGGCATTCACAACTTGT 958  
 Db 198 AGACAACCTTTCAGAAACTTTATGCAAAATTCAGTACAGCTTAATATGGACACCATGAAAT 257  
 Qy 959 GGTTCGCGGAGGAGTTTCATCCCTCTGGGCGAGCTTCAGCAAGCTCTCGGGGAGGGGCT 1018  
 Db 258 GCAGCTGGTGGGCTTTTCTCAGAGAGGAATTTCTCATGAGCTGTCTCGGAAAGTGT 317  
 Qy 1019 CCAGCAGCGCATGTTCTTCCTGGTTTCAACAGAGCTCCTGCTATACACAGCGCGGGGCTGAC 1078  
 Db 318 GCAACCTCGAATGTTTTCCTGTTTAAATGATGCGCCTCTGTATACACACAGTCGAGTC 377  
 Qy 1079 GGCCTCCCAATCAGTTTAA 1096  
 Db 378 TGGGATGTATAAACTGAA 395

RESULT 15  
 AAH15014  
 ID AAH15014 standard; cDNA; 1821 BP.  
 XX  
 AC AAH15014;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:12971.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 XX

thrombolic disease, polycystic ovarian syndrome, reduced fertility, complications of pregnancy, menstrual irregularities, hirsutism, stress incontinence and depression. The coding region of FGDL-like protein polynucleotide is useful for generating antisense oligonucleotides or ribozymes. These antisense oligonucleotides are useful for modulating FGDL-like protein gene expression. The FGDL-like protein is useful for generating antibodies against FGDL-like protein amino acid sequences and for use in various assay systems. The present sequence is the cDNA encoding the human FGDL-like protein.

Sequence 1296 BP; 415 A; 273 C; 281 G; 327 T; 0 other;

Query Match 7.3%; Score 83.6; DB 24; Length 1296; Best Local Similarity 51.3%; Pred. No. 1e-12; Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

Qy 719 TTACTACCGCTCAACACCTTCCTCCCTGGCGCCACTCCACGGCTCATGCACACTACAAGCA 778  
 Db 18 TAATCTGGCCCTCAAGCAGCTACCTGCTCAAGCCGGTTTCAGAGGATCCCCAGTACAGGCT 77  
 Qy 779 GGTCTCTGGAGCGCTGTGCACAAACACCCAGCCGCGGAGCTTCAGGGACTGCCCG 838  
 Db 78 GTTCTGCACAGATTATTTGAAGATCTCATAGAGATGCTGGAGATTACAGACACTCA 137  
 Qy 839 AGCCGCTTTGGCAGAGATCAGGAGATGGTGGCAGACTCCAGGGTACGATGATCAAGAT 898  
 Db 138 AGATGCCCTTCTGTTTATAGAGGTAGCAACACCGCCCAATGACACCATGAAGCAAGG 197  
 Qy 899 GGAGAAATTTCCAGAGCTGCACGAACTCAAGAAGATTGATTTGGCATTCACAACTTGT 958  
 Db 198 AGACAACCTTTCAGAAACTTTATGCAAAATTCAGTACAGCTTAATATGGACACCATGAAAT 257  
 Qy 959 GGTTCGCGGAGGAGTTTCATCCCTCTGGGCGAGCTTCAGCAAGCTCTCGGGGAGGGGCT 1018  
 Db 258 GCAGCTGGTGGGCTTTTCTCAGAGAGGAATTTCTCATGAGCTGTCTCGGAAAGTGT 317  
 Qy 1019 CCAGCAGCGCATGTTCTTCCTGGTTTCAACAGAGCTCCTGCTATACACAGCGCGGGGCTGAC 1078  
 Db 318 GCAACCTCGAATGTTTTCCTGTTTAAATGATGCGCCTCTGTATACACACAGTCGAGTC 377  
 Qy 1079 GGCCTCCCAATCAGTTTAA 1096  
 Db 378 TGGGATGTATAAACTGAA 395

RESULT 15  
 AAH15014  
 ID AAH15014 standard; cDNA; 1821 BP.  
 XX  
 AC AAH15014;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:12971.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 XX



GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:33:49 ; Search time 34.959 Seconds  
(without alignments)

10088.332 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_1352\_2501

Perfect score: 1150

Sequence: 1 cgggtacaacgagcgagc.....attgaggagagcggaagacga 1150

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2.6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.8	3.5	4776	US-08-852-401-1	Sequence 1, Appli
2	40	3.5	7218	US-08-232-463-14	Sequence 14, Appli
3	39.6	3.4	4722	US-08-979-608A-14	Sequence 14, Appli
4	38.4	3.3	4267	US-08-949-155-51	Sequence 51, Appli
5	38.4	3.3	4267	US-08-819-964-51	Sequence 51, Appli
6	37.2	3.2	44377	US-08-804-227C-7	Sequence 7, Appli
7	37.2	3.2	44377	US-08-804-198-1	Sequence 1, Appli
8	36.4	3.2	2824	US-09-010-928B-3	Sequence 3, Appli
9	36.2	3.1	4403765	US-09-103-840A-2	Sequence 2, Appli
10	36	3.1	2214	US-08-864-038A-1	Sequence 1, Appli
11	36	3.1	3331	US-08-864-038A-2	Sequence 2, Appli
12	36	3.1	3331	US-08-864-038A-4	Sequence 4, Appli
13	35.4	3.1	2132	US-09-552-322-1	Sequence 1, Appli
14	35.4	3.1	4137	US-09-499-984-2	Sequence 2, Appli
15	35.4	3.1	4403765	US-09-103-840A-2	Sequence 2, Appli
16	35.4	3.1	4411529	US-09-103-840A-1	Sequence 1, Appli
17	34.6	3.0	657	US-09-527-345-3	Sequence 3, Appli
18	34.6	3.0	1029	US-08-911-853-1	Sequence 1, Appli
19	34.6	3.0	1029	US-09-479-409-1	Sequence 1, Appli
20	34.6	3.0	1029	US-09-479-409-1	Sequence 1, Appli
21	34.6	3.0	4377	US-08-911-853-28	Sequence 28, Appli
22	34.6	3.0	4377	US-08-911-853-28	Sequence 28, Appli
23	34.6	3.0	4377	US-09-479-409-28	Sequence 28, Appli
24	34.6	3.0	10348	US-08-457-273B-41	Sequence 41, Appli
25	34.6	3.0	10348	US-08-457-273B-41	Sequence 41, Appli
26	34.6	3.0	10348	US-08-556-419-13	Sequence 13, Appli
27	34.6	3.0	10366	US-08-041-886-14	Sequence 14, Appli
				US-08-246-982A-5	Sequence 5, Appli

28	34.6	3.0	10366	1	US-08-453-265-5	Sequence 5, Appli
29	34.4	3.0	7042	4	US-09-092-508-1	Sequence 1, Appli
30	34.4	3.0	7042	4	US-09-435-115-1	Sequence 1, Appli
31	34.4	3.0	7042	4	US-09-098-310-1	Sequence 1, Appli
32	34.4	3.0	7042	4	US-09-690-364-21	Sequence 21, Appli
33	34.4	3.0	7075	4	US-09-092-508-15	Sequence 15, Appli
34	34.4	3.0	7075	4	US-09-435-115-15	Sequence 15, Appli
35	34.2	3.0	1331	4	US-09-335-983-14	Sequence 14, Appli
36	34.2	3.0	4195	1	US-08-340-011-1	Sequence 1, Appli
37	34.2	3.0	4195	3	US-08-901-710-1	Sequence 1, Appli
38	34.2	3.0	4416	3	US-08-795-430-1	Sequence 1, Appli
39	34.2	3.0	4416	4	US-09-355-700-1	Sequence 1, Appli
40	34.2	3.0	4416	4	US-08-601-132-36	Sequence 36, Appli
41	34.2	3.0	4425	1	US-08-222-616-31	Sequence 31, Appli
42	34.2	3.0	4425	4	US-08-446-648-31	Sequence 31, Appli
43	34.2	3.0	4425	5	PCT-US95-04228-31	Sequence 31, Appli
44	34.2	3.0	4795	1	US-08-340-011-3	Sequence 3, Appli
45	34.2	3.0	4795	3	US-08-901-710-3	Sequence 3, Appli

## ALIGNMENTS

RESULT 1  
US-08-852-401-1  
; Sequence 1, Application US/08852401  
; Patent No. 5976836  
; GENERAL INFORMATION:  
; APPLICANT: Weber, J. Mark  
; APPLICANT: Hessler, Paul E.  
; APPLICANT: Larsen, Peter E.  
; APPLICANT: Luu, B. Minh  
; TITLE OF INVENTION: Methods and Compositions for Enhancing  
; Erythromycin Production  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.  
; STREET: 2 Prudential Plaza, 180 N. Stetson, Suite  
; STREET: 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,401  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueller, Lisa L.  
; REGISTRATION NUMBER: 38,978  
; REFERENCE/DOCKET NUMBER: FER2159P00300S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5400  
; TELEFAX: 312-616-5460  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4776 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-852-401-1

Query Match 3.5%; Score 40.8; DB 2; Length 4776;  
Best Local Similarity 43.9%; Fred. No. 0.083;  
Matches 174; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 661 TCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTCGAGAGGTGTGTT 720

Db 797 TCAACCGCGCGGCGCATCAAGAGGACACCGCGCGCGTGTGCGCGCTCAACG 856  
Qy 721 ACCTACCGCTCAACACCTTCCTCTCGCGGCACCTCCACCGCGCTATGACACTACAGCAGG 780  
Db 857 AGATCGGCACACCGCGCGCGCATAGTGGCTGCGCGCACCGGGTAATCGCGCTGC 916  
Qy 781 TCCTGAGCGGTGTGCAAAACACACCGCGGAGCGCGGACTTTCAGGGACTGCCGAG 840  
Db 917 TGGTCCGAGAGTGTCAACCGGTTCCTCGCGCTTCGCGAGCGCTCGAGGCGCGCG 976  
Qy 841 CGCTTTGCGAGATACCGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGG 900  
Db 977 CGTCCGCGCGGCTACCGCTCGCTGCTGTCAACACCGCGCTCGGATGACGAGGAGG 1036  
Qy 901 AGAATTTCCAGAGCTGCACGAACACTCAAGAAGATTTGATGGCAATTCACAATCTTGTGG 960  
Db 1037 ACTAGTCGGATGCTATCGCGCGCGCGTGGAGGATGCTTCGTGTCGCGCGGAGA 1096  
Qy 961 TTCGGGAGGAGTTCATCGTCTGGGCGAGCTCAGCAAGCTCTCGGGGAGGGGCTCC 1020  
Db 1097 TCGCCACACCGAGGCGGAGAGCGGATCAGCGCGAGCTACTACAGAAGCTGTGGCGG 1156  
Qy 1021 AGCAGCGCATGTTCTTCTCTGTTCACAGCAGTCTCTGC 1056  
Db 1157 ACGGCGTGGCATGCTTCTCGTCAACGCGCGCGC 1192

## RESULT 2

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

; IMMEDIATE SOURCE:  
; CLONE: ptz9pt-Fls  
US-08-232-463-14

Query Match 3.5%; Score 40; DB 1; Length 7218;  
Best Local Similarity 5.4%; Pred. NO. 0.18;  
Matches 22; Conservative 206; Mismatches 176; Indels 0; Gaps 0;

Qy 280 AGGATGAGGCGGAGGAGAGATTCCCACTGATAAGCGTACTTCATAGTAAGGAG 339  
Db 1446 AGAATTTGGTACRR 1387  
Qy 340 TGTCTACCACGAGGAGAACATATCTGAAGATCTCGAAGTTATCATCTCGTGGTTTCAGA 399  
Db 1386 RRR 1327  
Qy 400 GCACAGTGAGCAAGAGGACGCCATCGCGAAGCACTGAAAGTCTCATATTCGCCGAAT 459  
Db 1326 RRR 1267  
Qy 460 TTGAACCTTTCACAAATTTTCATACTAATTTCTCAAGAAATTCAGCAACGACTGCC 519  
Db 1266 RRR 1207  
Qy 520 TGTGGGAGCGCGCTCAAAATGCCAAATCAGAGATTACCAAGAAATCGCGATGTCATGC 579  
Db 1206 RRR 1147  
Qy 580 TGAAGAACATTCAGGCGCATGAAGCACCTCGCGGCTCACTGTGGAAGCACACGAGGCT 639  
Db 1146 RRR 1087  
Qy 640 TGGAGCGCTCGAGAATGGAATCAGAGCTCCCGCGCGCTGGAG 683  
Db 1086 RRR 1043

## RESULT 3

US-08-979-608A-14  
; Sequence 14, Application US/08979608A  
; Patent No. 6355451  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; Law, Simon W.  
; Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
; TREATING ATHEROSCLEROSIS  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/979,608A  
; FILING DATE: 26-NOV-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/048,547  
; FILING DATE: 03-JUN-1997  
; APPLICATION NUMBER: US 60/031,930  
; FILING DATE: 27-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis  
; REGISTRATION NUMBER: 35,965



REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4722 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 61...1731  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-08-979-608A-14  
Query Match 3.4%; Score 39.6; DB 4; Length 4722;  
Best Local Similarity 47.5%; Pred. No. 0.18; Indels 6; Gaps 1;  
Matches 154; Conservative 0; Mismatches 164; Indels 6; Gaps 1;  
QY 599 GAAGCACCTGGCGGCTCACCTGTGAAGCACAGCGAGGCGCTTGGAGCCCTGGAGAAATGG 658  
Db 618 GGAGAGCTGGCGCTCTGTGCAAGAGTATGCGAACTCTCTGAGGAGCACCAGAACTC 677  
QY 659 AATCAAGAGCTCCGGGGCTGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTG 718  
Db 678 GCAGAAAGCAGATGAAGTGTCTGCAGAAAGAGCAGACGACTGTGTGAGGAGAGG--- 733  
QY 719 TTACCTACCGCTCAACACCTTCTCTCTGGCGCACCTGCACCGGCTCATGCACCTACAAGCA 778  
Db 734 --ACCACTCGTGGGAGCACAGCAAGGCGCATCTTGGCGCCGAGCAAGCTCGAGAGCCT 791  
QY 779 GGTCTGGAGCGCTGTGCAAAACACACCGCGAGCGGAGCTTCAAGGAGCTGCGCC 838  
Db 792 GTCCGGGAGCTCGAGCGGCACAAACCGCTCGCTCAAGGAAGAGTGTGCGAGCGAGCCG 851  
QY 839 AGCCGCTTTGGAGAGATCAGGAGATGTTGGCAGAGCTCCAGCTCCAGATGATCAAGAT 898  
Db 852 AGAGAGGAGGAGAGGAGCGGAAGAGGTGAGTCAACCTTCCAGATGACGCTCAACGACAT 911  
QY 899 GGAGAAATTCACAGAGCTGCACGA 922  
Db 912 TCAGTGCAGATGGAGCAGCACAA 935  
RESULT 4  
US-08-949-155-51  
Sequence 51, Application US/08949155  
Patent No. 6271436  
GENERAL INFORMATION:  
APPLICANT: Piedrahita, Jorge A  
APPLICANT: Bazer, Fuller W  
TITLE OF INVENTION: Compositions and Methods for the  
Generation of Transgenic Animal Species  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE AND DURKEE  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: US  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/949,155  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/027,338

FILING DATE: 11-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/046,094  
FILING DATE: 09-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Habler, David W.  
REGISTRATION NUMBER: 41,071  
REFERENCE/DOCKET NUMBER: TAMK:177  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4267 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-949-155-51  
Query Match 3.3%; Score 38.4; DB 4; Length 4267;  
Best Local Similarity 47.2%; Pred. No. 0.39;  
Matches 117; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
QY 24 GCCGCTCGCGGCCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 83  
Db 3020 GCCCTCTCTCGCGGCCACGAGGAGCTGATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 3079  
QY 84 AAACCTCAGCCCCCGCAGCGCAACGACAGGCTCCCTGACTGGCAGTCTCACCCTTCCGAG 143  
Db 3080 GAGGAGCTGGAGCGCGAGGCTGGGCGCCGCTGACCCAGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 3139  
QY 144 CTGCTCTGAACTCGCAGGGGGAGTGGCCCTCGCCAAAGTGAACCTTGTCTCCCAACCTG 203  
Db 3140 GAGCTGCAGCGCGCAGCGCCCGCTGGGCGCGAGCATGGAGGAGCTGCGCAACCGCTTG 3199  
QY 204 AGCCCGCAGCACCAAGCAGGCTCTCCCTTGATCAGCCCGCTGCTGATGACAGGAGGCTG 263  
Db 3200 GTGCTCTACCGCAGGAGGTGCAACATGTTGGCGCACCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 3259  
QY 264 CCCCAGGAC 271  
Db 3260 CGCCTGGC 3267  
RESULT 5  
US-09-819-964-51  
Sequence 51, Application US/09819964  
Patent No. 6369294  
GENERAL INFORMATION:  
APPLICANT: Piedrahita, Jorge A  
APPLICANT: Bazer, Fuller W  
TITLE OF INVENTION: Compositions and Methods for the  
Generation of Transgenic Animal Species  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE AND DURKEE  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: US  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/819,964  
FILING DATE: 28-Mar-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/949,155  
FILING DATE: <Unknown>



```

ADDRESS: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

```

Query Match	3.28;	Score	37.2;	DB	2;	Length	44377;
Best Local Similarity	43.89;	Pred.	No.	2.9;			
Matches	162;	Conservative	0;	Mismatches	208;	Indels	0;
Gaps	0;						
Qy	528	GGCCGCTCAAAATGCCAAATCAGAGATTACCAAGAATCGCGGATGTCATGCTGAAGAAC	587				
Db	29515	GACTTCGGGACCGGCTTCCACCGAGGACGGCAGCTCCCGCGGGAGCTGCTCTCGAAC	29574				
Qy	588	ATTGAGGCGATGAAGCACCTTCGCGGCTCACCTGTGTGAAGACACAGCGAGCCCTTGGAGGCC	647				
Db	29575	TCGCTCACCAAGGAGTTCGTGAGCGCTCCCTCGGCTGCTCCGTCGGGCGCGCGGCTTC	29634				
Qy	648	CTGAGAAATGGAATCAAGAGCTCCCGGGGCTTGGAACATTTCTGCAGAGACTTTGAGCTG	707				
Db	29635	CTGAGACTGGGCAAGACCGAGCTCGGACCCCGAGCGGATCGCCGCGGAACACCCCGGG	29694				
Qy	708	CAGAAGGTGTGTTACCTACCGCTCAACACCTTCCTCTCTCGCGGCCACTGCACCGGCTCATG	767				
Db	29695	GTGGCTTACCGGGGCGTTTCGAGCTCAACGAGCGCGGACCCGACGCTCGGCGGGCTGCTG	29754				
Qy	768	CACTAACAAGCAGGTCTTGAGCGGCTGTGTCAAAACACCAACCCCGCGGAGCCAGCCGACTTC	827				
Db	29755	CGGGAACGTATGGACCTGTTCCGCGCGGCGTGTCTGCACCCGCTGCCGCTGCTACCCAC	29814				
Qy	828	AGGACTTGCAGCGCGCTTTTGGCAGAGATCACGGAGATGTTGGCACAGCTCCACGGTACG	887				
Db	29815	GACGTGGCGCGGGCGGAGCGGCTCGGCACCATCAGCCAGGGCGGACACCGGAAG	29874				

Qy 888 ATGATCAAGA 897  
     |   |   |  
 Db 29875 CTCGCTCTGA 29884  
  
 RESULT 8  
 US-09-010-928B-3/c  
 ; Sequence 3, Application US/09010928B  
 ; Patent No. 5994099  
 ; GENERAL INFORMATION:  
 ;   APPLICANT: Lewis, Randolph V  
 ;   APPLICANT: Hayashi, Cheryl Y  
 ;   TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK  
 ;   TITLE OF INVENTION: CODING THEREFOR  
 ;   NUMBER OF SEQUENCES: 29  
 ;   CORRESPONDENCE ADDRESS:  
 ;   ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
 ;   STREET: 8110 GATEHOUSE RD. SUITE 500E  
 ;   CITY: FALLS CHURCH  
 ;   STATE: VIRGINIA  
 ;   COUNTRY: UNITED STATES OF AMERICA  
 ;   ZIP: 22042  
 ;   COMPUTER READABLE FORM:  
     MEDIUM TYPE: Floppy disk  
     COMPUTER: IBM PC compatible  
     OPERATING SYSTEM: PC-DOS/MS-DOS  
     SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ;   CURRENT APPLICATION DATA:  
     APPLICATION NUMBER: US/09/010,928B  
     FILING DATE: 22-JAN-1998  
     CLASSIFICATION: 435  
 ;   ATTORNEY/AGENT INFORMATION:  
     NAME: Murphy Jr., Gerald M  
     REGISTRATION NUMBER: 28977  
     REFERENCE/DOCKET NUMBER: 1447-109P  
 ;   INFORMATION FOR SEQ ID NO: 3:  
     SEQUENCE CHARACTERISTICS:  
         LENGTH: 2824 base pairs  
         TYPE: nucleic acid  
         STRANDEDNESS: not relevant  
         TOPOLOGY: linear  
         MOLECULE TYPE: cDNA  
         HYPOTHETICAL: NO  
         FEATURE:  
             NAME/KEY: -  
             LOCATION: 1..2824  
             OTHER INFORMATION: /note= "Flagelliform DNA se  
             OTHER INFORMATION: taken from 3' region. Stop  
             FEATURE:  
                 NAME/KEY: CDS  
                 LOCATION: 1..2724  
 ; US-09-010-928B-3

Query Match	3.2%;	Score 36.4;	DB 2:	Length 2824;
Best Local Similarity	48.5%;	Prod. No. 1.2;		
Matches 100;	Conservative	0;	Mismatches 106;	Indels 0; Gaps 0;
Qy 721	ACCTACGCGCTCAACAGCTTCCTCGTGGCGGCACATGCACGGCTCATGCATACAAACAGG	780		
Db 678	ACCTCTCTGCTCGCGCTGCTCTCCAGCGCCACCAGCACCCAGGTCCTCACCAGG	619		
Qy 781	TCCTGGAGCGGCTGTGCAAAACACCAACCCCGCGAGCCACGCCGACTTCAGGGACTCCCGAG	840		
Db 618	TCCAGCACCAACAGCTCCAGCTCCACCAGGTCCACGACGACGAGTCCAGGTCACCAGG	559		
Qy 841	CCGCTTTTGGCAGAGATACGGAGATGGTGGGCACAGCTCCACGGTACGATGATCAAGATGG	900		
Db 558	TCCAGCACCAACAGGTCAGCACCCACAGGTCCAGCACCAACAGGTCCTCACCAGG	499		
Qy 901	AGAAATTTCCAGAAGCTGCACGAATCTC	926		
Db 498	TCCAGCACCAACAGGTCCTCCTCAAAACAC	473		

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2214
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; US-08-864-038A-1

Query Match 3.1%; Score 36; DB 3; Length 2214;
Best Local Similarity 55.6%; Pred. No. 1.4; Indels 0; Gaps 0;
Matches 69; Conservative 0; Mismatches 55;

QY 704 GCTGCGAAGGTGTGTACCTACGGCTCAACACACTTCTCCTGTGGCGCACTGCACC GGCT 763
      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 1073 GCTGCACGGCAGGGGAGCACAGCGCCCTCCGCCCTCTCTCCACCACTTCACCTCCT 1014

QY 764 CATGCATACAAGCAGTGCTCTGGAGCGGTGTGCAAAACACACCCGCCGAGCCACGCCGA 823
      | ||| ||| | | | | | | | | | | | | | | | | | | | |
Db 1013 CGCGCACCAACGCGGTGTGGGGGGGAGCTGCAGCAGCGCAGCAGCAGCTCTGCGG 954

QY 824 CTTC 827
      |||
Db 953 CTC 950

RESULT 11
US-08-864-038A-2/C
; Sequence 2, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
```

[illegible]



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 18:25:06 ; Search time 37.9989 Seconds  
(without alignments)  
11655.134 Million cell updates/sec

Title: US-09-555-342b-1\_COPY\_1352\_2501

Perfect score: 1150

Sequence: 1 cgggtacaagcagcgacgac.....attgaggagagcgaagacga 1150

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	762	66.3	3094	10	US-09-925-297-220
2	297	25.8	379	10	US-09-960-352-154
3	292	25.4	400	10	US-09-960-352-156
4	248.6	21.6	408	10	US-09-783-590-3620
5	237.4	20.6	500	10	US-09-783-590-3621
6	190.8	16.6	335	10	US-09-783-590-3600
7	168.8	14.7	356	10	US-09-783-590-3575
8	154.6	13.4	426	10	US-09-815-343-1490
9	133.2	11.6	288	10	US-09-783-590-3542
10	106	9.2	1718	9	US-09-764-868-51
11	105.2	9.1	716	9	US-09-764-868-475
12	64.4	5.6	2686	9	US-09-764-868-48
13	64.4	5.6	3380	10	US-09-799-799-1
14	64	5.6	1091	9	US-09-764-868-316
15	43.4	3.8	3786	10	US-09-815-242-7865
16	42.4	3.7	1730	9	US-09-860-670-78
17	42.4	3.7	1730	9	US-09-764-868-52
18	42.4	3.7	1730	9	US-09-764-868-476
19	40.8	3.5	2172	10	US-09-815-242-4038

20	40.4	3.5	598	10	US-09-770-149-991	Sequence 991, App
21	40.4	3.5	768	9	US-09-938-842A-812	Sequence 812, App
22	39.6	3.4	4722	10	US-09-962-055-14	Sequence 14, Appl
23	39.6	3.4	4722	12	US-10-023-529-14	Sequence 14, Appl
24	39.6	3.4	4722	12	US-10-023-523-14	Sequence 14, Appl
25	39.4	3.4	420	10	US-09-960-352-9773	Sequence 9773, Ap
26	38.4	3.3	1685	9	US-09-954-531-133	Sequence 133, App
27	38.4	3.3	1685	9	US-09-954-531-354	Sequence 354, App
28	38.4	3.3	1685	10	US-09-962-436-261	Sequence 261, App
29	38.4	3.3	2296	10	US-09-822-849A-259	Sequence 259, App
30	37.8	3.3	422	10	US-09-960-352-9532	Sequence 9532, Ap
31	37.8	3.3	422	10	US-09-738-973-337	Sequence 337, App
32	37.4	3.3	552	9	US-09-764-868-406	Sequence 406, App
33	37.4	3.3	16747	10	US-09-764-877-3354	Sequence 3354, Ap
34	36.6	3.2	273	10	US-09-864-761-17066	Sequence 17066, A
35	36.6	3.2	403	10	US-09-960-352-11685	Sequence 11685, A
36	36.6	3.2	1369	10	US-09-443-704-17	Sequence 17, Appl
37	36.2	3.1	866	10	US-09-764-870-262	Sequence 262, App
38	36.2	3.1	3121	9	US-10-033-245-6	Sequence 6, Appl
39	36.2	3.1	3121	12	US-10-033-223-6	Sequence 6, Appl
40	36.2	3.1	3121	12	US-10-033-246-6	Sequence 6, Appl
41	36.2	3.1	3121	12	US-10-033-301-6	Sequence 6, Appl
42	36.2	3.1	3121	12	US-10-033-326-6	Sequence 6, Appl
43	36	3.1	1100	10	US-09-263-959-271	Sequence 271, App
44	36	3.1	4966	10	US-09-764-869-1885	Sequence 1885, Ap
45	36	3.1	7238	10	US-09-764-869-1886	Sequence 1886, Ap

#### ALIGNMENTS

RESULT 1  
US-09-925-297-220  
; Sequence 220 Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 220  
; LENGTH: 3094  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (4)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-297-220

Query Match 66.3%; Score 762; DB 10; Length 3094;					
Best Local Similarity 100.0%; Pred. No. 2.4e-209;					
Matches 762; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	389	GTGCTTTCAGACACAGTCAGCAAGAGAGCGCCATGCCGGAAGCACTGAAAGTCTCAT	448		
Db	62	GTGGTTTCAGACACAGTCAGCAAGAGAGCGCCATGCCGGAAGCACTGAAAGTCTCAT	121		
QY	449	ATTCCCGAATTTTGACCTTTGCACAAATTTTCTACTAATTTTCTCAAGAAATTCAGCA	508		
Db	122	ATTCCCGAATTTTGACCTTTGCACAAATTTTCTACTAATTTTCTCAAGAAATTCAGCA	181		
QY	509	ACGACTTGCCCTGTGGGAAGCGCGCTCAAAATGCCCAATACAGAGATTACCAAAATTCGG	568		



```
Db 182 AGCACTTCCCTGTGGAGGCGCTCAAAATGCCAAATCAGAGATTACCAAGAATCGG 241
Qy 569 CGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTACCTGTGGAACA 628
Db 242 CGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTACCTGTGGAACA 301
Qy 629 CAGCGAGGCGCTTGGAGGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 688
Db 302 CAGCGAGGCGCTTGGAGGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 361
Qy 689 CTGCAGAGACTTGAAGTGCAGAGGCTGTGTACCTACCGCTCAACACTTCTCCTCGG 748
Db 362 CTGCAGAGACTTGAAGTGCAGAGGCTGTGTACCTACCGCTCAACACTTCTCCTCGG 421
Qy 749 GCACTGACCGGCTTCATGCACTACAAGCAGTCTGAGCGGCTGTCACAAACACCC 808
Db 422 GCACTGACCGGCTTCATGCACTACAAGCAGTCTGAGCGGCTGTCACAAACACCC 481
Qy 809 GCCGAGCCACGCGGCTTCAGGAGCTGCGGAGCGCTTTGGCAGAGATCAGGAGATGGT 868
Db 482 GCCGAGCCACGCGGCTTCAGGAGCTGCGGAGCGCTTTGGCAGAGATCAGGAGATGGT 541
Qy 869 GCCAGAGCTCCAGGTACAGATCAAGATGAGAAATTCAGAAAGCTGCAGAACTCAA 928
Db 542 GCCAGAGCTCCAGGTACAGATCAAGATGAGAAATTCAGAAAGCTGCAGAACTCAA 601
Qy 929 GAAAGATTGATTGCATTCGACAAATCTGTGGTCCGGAAGGAGTTCATCCGCTCGG 988
Db 602 GAAAGATTGATTGCATTCGACAAATCTGTGGTCCGGAAGGAGTTCATCCGCTCGG 661
Qy 989 CAGCCTCAGCAAGCTCTCGGGAGGGGCTCCAGCAGCGCATGTTCTTCTCTTCAACGA 1048
Db 662 CAGCCTCAGCAAGCTCTCGGGAGGGGCTCCAGCAGCGCATGTTCTTCTTCAACGA 721
Qy 1049 CGTCTCTATACAGAGCGGGGCTGAGCGCTTCCAAATCAATGTTAAAGTCCAGGGCA 1108
Db 722 CGTCTCTATACAGAGCGGGGCTGAGCGCTTCCAAATCAATGTTAAAGTCCAGGGCA 781
Qy 1109 GCTCCGCTCTATGCATGACCATTCAGGAGAGCGAAAGCA 1150
Db 782 GCTCCGCTCTATGCATGACCATTCAGGAGAGCGAAAGCA 823

RESULT 2
US-09-960-352-154
; Sequence 154, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 154
; LENGTH: 379
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E1-A9
US-09-960-352-154

Query Match 25.8%; Score 297; DB 10; Length 379;
Best Local Similarity 89.7%; Pred. No. 5.7e-76;
Matches 331; Conservative 0; Mismatches 35; Indels 3; Gaps 1;

Qy 284 TGAGGCGCGGAGAGAGATCCCACTCATAAAGGCTACTCATAGCTTAAGGAAGTCTC 343
Db 11 TGGGCGCGGAGAGAGATCCCACTCATAAAGGCTACTCATAGCTTAAGGAAGTCTC 70
```

```
Qy 344 TACCACCGAGCGAACATATCTGAAGGATCTCGAAAGTTATCACTTCGTGGTTTCAGAGCAC 403
Db 71 CACCACCTGAGAGGACATATCTGAAGGATCTTGAAGTCACTACTTCGTGGTTTCAGAGCAC 130
Qy 404 AGTGAGCAAGAGAGAGCGGATCCCGGAAGCACTGAAAGTCTCATATATCCCGAAATTTGA 463
Db 131 AGTGAGCAAGAGAGGACTCCATGCCGGAACCTTGAAGAGTCTCATATATCCCGAAATTTGA 190
Qy 464 ACCTTTGACAAATTTTCATACATAATTTTCAAGGAAATTTGAGCAAGCACTTGCCTCTGTG 523
Db 191 ACCTTTGACAAATTTTCACAAATTTTCTCAAGGAAATTTGAGCAAGCACTTGCCTCTGTG 250
Qy 524 GGAAGCGCGCTCAATGCCCAATCA---GAGATTACCAAGAAATTCGCGATGTGATGCT 580
Db 251 GGAAGCGCGCTCGAATGCCACATCAGAGGAGATTTACCAGAGAAATCGGAGATGTAACTCT 310
Qy 581 GAAGAACTTCAGGCGATGAAGCACTGGCGGCTCACTGTGGAAGCAAGCGAGGCGCTT 640
Db 311 GAAGAACTTCAGGCGATGAAGCACTGGCGGCGCACTGTGGAAGCAAGCGAGGCGCT 370
Qy 641 GGAGCGCCT 649
Db 371 GGAGCGCCT 379

RESULT 3
US-09-960-352-156
; Sequence 156, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 156
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E2-A9
US-09-960-352-156

Query Match 25.4%; Score 292; DB 10; Length 400;
Best Local Similarity 87.5%; Pred. No. 1.6e-74;
Matches 343; Conservative 0; Mismatches 45; Indels 4; Gaps 2;

Qy 284 TGAGGCGCGGAGAGAGATCCCACTCATAAAGGCTACTCATAGCTTAAGGAAGTCTC 343
Db 9 TGGGCGCGGAGAGAGATCCCACTCATAAAGGCTACTCATAGCTTAAGGAAGTCTC 68
Qy 344 TACCACCGAGCGAACATATCTGAAGGATCTCGAAAGTTATCACTTCGTGGTTTCAGAGCAC 403
Db 69 CACCACCTGAGAGGACATATCTGAAGGATCTTGAAGTCACTACTTCGTGGTTTCAGAGCAC 128
Qy 404 AGTGAGCAAGAGAGGCGGATCCCGGAAGCACTGAAAGTCTCATATATCCCGAAATTTGA 463
Db 129 AGTGAGCAAGAGAGGACTCCATGCCGGAACCTTGAAGAGTCTCATATATCCCGAAATTTGA 188
Qy 464 ACCTTTGACAAATTTTCATACATAATTTTCAAGGAAATTTGAGCAAGCACTTGCCTCTGTG 523
Db 189 ACCTTTGACAAATTTTCACAAATTTTCTCAAGGAAATTTGAGCAAGCACTTGCCTCTGTG 248
Qy 524 GGAAGCGCGCTCAATGCCCAATCA---GAGATTACCAAGAAATTCGCGATGTGATGCT 580
Db 249 GGAAGCGCGCTCGAATGCCACATCAGAGGAGATTTACCAGAAATCGGAGATGTAGTCT 308
Qy 581 GAAGAACTTCAGGCGATGAAGCACTGTG---GCGGCTCACTGTGGAAGCAAGCGAGGCGCT 639
```



OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (413)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (436)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (450)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (455)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (459)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (461)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (475)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (477)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (487)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (499)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-3621

Query Match 20.6%; Score 237.4; DB 10; Length 500;  
Best Local Similarity 83.3%; Pred. No. 9.4e-59;  
Matches 415; Conservative 0; Mismatches 60; Indels 23; Gaps 13:  
QY 420 GCATGCGCGGAGCACTGAAAAGTCTCATATTCCTCGAATTTTGAACCTTTGCACAAATTT 479  
Db 1 GCAGAGCGGAGCACTGAAAAGTCTCATATTCCTCGAATTTTGAACCTTTGCACAAATTT 60  
QY 480 CATACTAATTTTCTCAAGGAAATGAGCAACGACTTGGCCCTGT-GGGAGGCGGCTCAAA 538  
Db 61 CATACTAATTTTNTCAAGGAAATGAGCAACGACTTGGCCCTGTGGGAGGCGGCTCAAA 120  
QY 539 TGCCCAATACAGATTACCAAGATCGCGGATGTCATGCTGAA-GAACATTCAGGCA 597  
Db 121 TGCCCAATACAGATTACCAAGANTCGCGGATGTCATGCTGAAAGGAACATTCAGGCA 180  
QY 598 TGAAGCACCTGGCGG-CTCACCTGT-GGAAGCACAGCG-AGGCCCTGGAGGCCCTGGAGA 654  
Db 181 TGAAGCACCTGGCGGTCTCACCTGTGGGAGCACAGGAGGCCCTTGGAGGCCCTGGAGN 240  
QY 655 A--TGAATCAAGAGTCCCGCGG--CTGGAGAACTTCTGCAGAGACTTT-GAGCTGCA 709  
Db 241 AATGGNATTCAGAGTCCCGCGGCTGGAGGANTTCTGCAGAGACTTTGGAGTGA 300  
QY 710 GAAGGTGTG-TTACCTACCGCTCAACACTT--CCTCTGCGGCGCACTGCACCGGCTCAT 766  
Db 301 GAAGGTGTGTTTACCTACCGCTCAACACTTTCCTTCTGCGGCGCACTGCACCGGTTTAT 360  
QY 767 GCACCTACAAG--CAGTCTCTGGAGGGCTGTGCAAAACACACCCCGGAGCCACCGCGA- 823  
Db 361 GCACCTACAAGCGAGTCTGAGGCGGNTTTCGAAAACACACCCCGATCGNGGCCAAG 420  
QY 824 -----CTTCAGGAGTCCGCGCGCTTTGGCAGAGATTCACGGAGATGGTGGCACAGCT 877  
Db 421 GCGGATTTTCAGGANTGCCGAGCGG-TTTCAGAGGTTNANGAGTTCTTTGGGANANTT 479  
QY 878 CCACGGTACGATGCA 895  
Db 480 CCAGGGTNGTGTTC 497

RESULT 6  
US-09-783-590-3600  
Sequence 3600, Application US/09783590  
Patent No. US20020110850A1  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick J.  
APPLICANT: Haseltine, William A.  
APPLICANT: Li, Haodong  
APPLICANT: Rosen, Craig A.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
FILE REFERENCE: PO-16.2C1  
CURRENT APPLICATION NUMBER: US/09/783,590  
CURRENT FILING DATE: 2000-02-15  
PRIOR APPLICATION NUMBER: 08/420,856  
PRIOR FILING DATE: 1995-04-12  
PRIOR APPLICATION NUMBER: 08/346,731  
PRIOR FILING DATE: 1994-11-21  
NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3600  
LENGTH: 335  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (5)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (29)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (36)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (42)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (72)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (73)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (144)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (157)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (184)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (246)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (254)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (271)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (278)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (298)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (329)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (333)

```
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3600

Query Match      16.6%; Score 190.8; DB 10; Length 335;
Best Local Similarity 84.0%; Pred. No. 2e-45;
Matches 278; Conservative 0; Mismatches 42; Indels 11; Gaps 6;

QY 420 GCCATGCCGGAAGCACTGAAAAGTCTCATATCCCGAATTTTGAACCTTTGCAACAATTT 479
Db 1 GGCANAGCGGAAGCACTGAAAAGTCTCANATCCNAAATTTNAAACCTTTGCAACAATTT 60

QY 480 CATACTAAATTTTCTCAAGGAAATGAGCAACGACTTCCCTGTGGGAAGCCGCTCAAT 539
Db 61 AATACTAAATTTTNTCAAGGAAATTAAGCAACGACTTCCCTGTGGGAAGCCGCTCAAT 120

QY 540 GCCCAATCAG-AGATACCAAGAAATCGGCGATGTCATGCTGAAGAACAATTCAGGCA 597
Db 121 GCCCAATCAGAGATTAACCAANAAATCGCGATGTCATGCTGAAGAACAATTCAGGGCA 180

QY 598 TGAAGCACCTGGCGGCTCACTCTGGAAGCACAGCGAGGCGCTTGGAGGCCCTGGAGAA-- 655
Db 181 TGANGCACCTGGGCGCTCACTCTGGAAGCACAGCGAGGCGCTTGGGAAGCCCTGGAAG 240

QY 656 ---TGGAAACAAGAGCTCCCGGC-GGCTGGAGAACTTCTCAGA-GACTTTGAGCTGC-- 708
Db 241 ATTGGNATTCAGNGCTCCCGCGCGGCTGGGAACCTTNTGCAGAGGACTTTGAGCTGCA 300

QY 709 AGAAGGTGTGTTACCTACCGCTCAACACCTT 739
Db 301 GAAAGGTGTGTTACCTACCGCTTCAAAAANTT 331

RESULT 7
US-09-783-590-3575
; Sequence 3575, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3575
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (32)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (45)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (51)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (61)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
```

```
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (75)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (134)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (145)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (168)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (216)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (246)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (267)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (276)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (280)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (325)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (335)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (338)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3575

Query Match      14.7%; Score 168.8; DB 10; Length 356;
Best Local Similarity 77.9%; Pred. No. 4.4e-39;
Matches 247; Conservative 0; Mismatches 57; Indels 13; Gaps 4;

QY 420 GCCATGCCGGAAGCACTGAAAAGTCTCATATCCCGAATTTTGAACCTTTGCAACAATTT 479
Db 1 GGCAGAGCGGAAGCACTGAAAAGTCTCATATCCCGAATTTTNAANCCTTTNCACAAATTT 60

QY 480 CATACTAAATTTTCTCAAGGAAATGAGCAACGACTTCCCTGTGGGAAGCCGCTCAAT 539
Db 61 NATACTAAATTTCTNTNAAAGGAAATTAAGCAACGACTTGCCTGTGGGAAGCCGCTCAAT 120

QY 540 GCCCAATCAGAGATTAACCAAGAAATCGGCGATGT-CATGCTGAAGAACAATTCAGGGCAT 598
Db 121 GCCCAATCAGAGNTTACCNAAGANTCGGGGATGTCATGCTGAAGANCATTCAGGGCAT 180

QY 599 GAAGCACCTGGCGGCTCACCTGT-GGAAGCACAGCGAGGCGCTTGGAGGCCCTGGAGAATG 657
Db 181 GAGGCACCTGGGGGTTCACTGTGGGAAGCACAGCNAGGCGCTTGGGAAGCCCTGGAAG 240

QY 658 GAATCAA-----GAGCTCCCGCGGCTGGAGAACTCTG-----CAGAGACTTTGAGCT 706
Db 241 AATGGNATTCAGAGGTTCCCGCGGNTGGAGGAATTTNTGCCAGAGAACTTTGAAGCT 300

QY 707 GCAGAGGTGTGTTACC 723
Db 301 GCAGAGGTGTGTTAAC 317

RESULT 8
```

```

US-09-815-343-1490/C
; Sequence 1490, Application US/09815343
; Patent No. US2001005596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun E.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1490
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-1490

Query Match      13.4%; Score 154.6; DB 10; Length 426;
Best Local Similarity 61.9%; Pred. No. 5.9e-35;
Matches 263; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

QY  321 TACTTCATAGTAAGGAAGTCTCTACCACCAGGCGGAACATATCTGAAGATCTCGAAGTT 380
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   426 TACTTCATAGTCAAGAGATCTCGCTACAGAAGCAACATACCTCAAGGATTTAGAACTT 367

QY  381 ATCACTTCGTGGTTTCAGAGCACAGTGAGCAAAAGAGGCGCATGCCGGAAGCACTGAAA 440
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   366 ATTACCGTGTGGTTCCCGACGCGCAGTGGTGAAGGAGGAGCGCATGCCCTGCAGCTCTGATG 307

QY  441 AGTCATCATATCCCGAATTTTGAACCTTTTGACAAATTTTCATCTAATTTTCTCAAGGAA 500
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   306 ACCTGTCTCTCTCCACATCATGCCATCTATGAGTTCACAGAGGCTTCTTGGCGCAG 247

QY  501 ATTGAGCAAGACATTTGCCCTGTGGGAAGGCGCTCAATATGCCAAATCAGAG---ATTAC 557
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   246 GTGGAGCAGAGGCTGGCAGCTCTGGGAAGGGCCCTCCAAAGCCCAACAAAAGGCGAGTCAT 187

QY  558 CAAGAATTCGGCATGTCTCATGAGCAACATTCAGGGCATCAGGCACCTGGCGCTCAC 617
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   186 CAACGAATCGGGACATCTCTGCTCAGGAACATGCGCCAGTTTAAAGGAGTTTACCAGCTAC 127

QY  618 CTGTGGAAGCACACGCGGCCCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGCGCG 677
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   126 TTCCAAGACATGACGAGATCTTAACAGACTGGAAGAGCTACCAACGCTGTGAAGAG 67

QY  678 CTGGAGAACTTCTCGAGAGACTTTGAGTCGAGAAGGTGTGTTACCTACCGCTCAACACC 737
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   66 TTGGAGCAGTGTAACAAGGAGTTTGAGCTGAGAAGGTCTGCTACTTGCCTCTCAACACG 7

QY  738 TTCCT 742
      ||| |||
Db   6 TTCCT 2

RESULT 9
US-09-783-590-3542
; Sequence 3542, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12

```

```

: PRIOR APPLICATION NUMBER: 08/346,731
: PRIOR FILING DATE: 1994-11-21
: NUMBER OF SEQ ID NOS: 12485
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3542
: LENGTH: 288
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (3)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (42)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (72)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (75)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (93)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (129)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (142)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (155)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (156)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (166)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (168)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (182)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (253)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (263)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (264)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (267)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (270)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (272)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (282)
: OTHER INFORMATION: n equals a,t,g, or c
: US-09-783-590-3542
:
: Query Match 11.6%; Score 133
: Best Local Similarity 85.8%; Pred. No.
: Matches 188; Conservative 0; Mismatch
:
: 427 CGGAAGCACTGAAAGTCTCATATTCGCCGAATTC

```

Query Match 11.6%; Score 133.2; DB 10; Length 288;  
Best Local Similarity 85.8%; Pred. No. 7e-29;  
Matches 188; Conservative 0; Mismatches 27; Indels 4;  
427 CGGAGCACTGAAAAGCTCTCATATCCCGAAATTTTGAACCTTTGCAAAATTTTCAAT

Db	8	CGGAGCAGCTGAAAGGCTCAAAATTCCTCCAAATTTAAACCTTTGACAAATTTAAAACTA	67
Qy	487	ATTTTCTCAAGGAAATTGAGCACGACTTGCCTCTGGGAGGCGCTCAAAATGCCCAA	546
Db	68	ATTNTNTAAGGAAATTAAGCAACGCTTTGCCCTCTGGAAGGCGCTCAAAATGCCCAA	127
Qy	547	TCAGAGATTACCAAGAAATCGCGGATGT-CATGCTGAAG-AACATTTCAGGCGCATGAAGCA	604
Db	128	TNA-AGATTACCAANAATCGCGGATGTNNATGCTGAAGNANCATTTCAGGCGATGGCA	186
Qy	605	CCTGGCGC-TCACCTGTGGAAGCACAGCGAGGCGCTTG	642
Db	187	CCTGGCGGCTTCACCTGTGGGAAGCACGCGAGGCGCTTG	225

```

RESULT 10
US-09-764-868-51
; Sequence 51, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-51

```

Query Match	9.28;	Score 106;	DB 9;	Length 1718;
Best Local Similarity	64.96;	Pred. No. 1.2e-20;		
Matches 157;	Conservative 0;	Mismatches 85;	Indels 0;	Gaps 0;
Qy	909	CAGAAGCTGCACGAACTCAAGAAAGATTGATTGGCATTTGACATCTGTGTGTTCCGGGA	968	
Db	13	CAGAAGCTAACGGAGCTGCACGGGACCTGCTGGCATAGAACTCATTTGCTCTGC	72	
Qy	969	AGGAGATTATCCCTCTGGGCAGGCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGGCG	1028	
Db	73	AGGAGATTATCCCTGAGGGCTGCTTCAAGACTCACCAGAAGAGGGCTCGACGAGAG	132	
Qy	1029	ATGTTCTTCTGTTCACGAGCTCCTGTATPACACGAGCCGGGGCTCACGGCTCTCAAT	1088	
Db	133	ATGTTTTTCTGTCTCAGATATGTTGCTGTACACAAGCAAGGAGTTGCAGGGACCAGC	192	
Qy	1089	CAGTTTTAAAGTCCACGGCAGCTCCGCTCTATGGCATGACATTGAGGAGAGGGAAGAC	1148	
Db	193	CAC'TCCGGATCCGGGGCTCTCTTCCCTCCAAAGCATGCTGGTGGAGAAAGTGATAAC	252	
Qy	1149	GA 1150		
Db	253	GA 254		

```

RESULT 11
US-09-764-868-475
; Sequence 475, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ- ID NO 475
;
; LENGTH: 716
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: SITE
;
; LOCATION: (715)
;
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-868-475

```

Query Match	9.1%	Score 105.2;	DB 9;	Length 716;
Best Local Similarity	64.0%;	Pred. No. 1.3e-20;		
Matches 155;	Conservative 2;	Mismatches 85;	Indels 0;	Gaps 0;
Qy	909	CAGAACTGCACGAACCTCAAGAAGATTGATTTGGCATTTGACAATCTTGTGGTTCCGGGA	968	
Db		::		
	3	CAGAAGCTWACGGAGCTCWRKCGGACCTGGTGGS CATAGAGAACCTCATTTGCTTCCTGGC	62	
Qy	969	AGGAGATTCAATCCGTC TGGCGAGGCTCAGCAAGCTCTCGGGGAAGGGGCTCCACAGCGC	1028	
Db	63	AGGAGATTCAATCCGTTGAGGGCTGCC TTCACAAAGCTCACAAAGAGGGGCTGCAGCAGAGG	122	
Qy	1029	ATGTTCTTCCTGTTCAACGAGCTCTGTATACACGAGCGGGGCTCAGCGGCTCCAAT	1088	
Db	123	ATGTTTTTCTGTTCTCAGATATGTTGCTGTACACAACAAAGGATTGCAGGGACCAGC	182	
Qy	1089	CAGTTTAAGTCCACGGGCAGCTCCCGCTCTATGGCATTGACGATTGAGGAGAGCGAAGAC	1148	
Db	183	CACITCCGGATCCGGGGCTCTCTCCCCCTCCAAGGCATGCTGGTGGAGAAGATGATAAC	242	
Qy	1149	GA 1150		
Db	243	GA 244		

```

RESULT 12
US-09-764-868-48
: Sequence 48, Application US/09764868
: Patent No. US20020168711A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PTZ32
: CURRENT APPLICATION NUMBER: US/09/764,868
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1510
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 48
: LENGTH: 2686
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (2505)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (2569)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (2644)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (2645)
: OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-48

```

```
Query Match      5.6%; Score 64.4; DB 9; Length 2686;
Best Local Similarity 47.8%; Pred. No. 1.5e-08;
Matches 287; Conservative 0; Mismatches 301; Indels 12; Gaps 3;

Qy 556 ACCAAGAANTCGGCGATGTCTAGTGAAGACATTTCAGGGCATTAAGCACCTTGCGGC 615
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 1505 ACCACGGCTCGGGACATCTCGAGAAGCTGGCCCATCTCTGAAAGATGTACGGCGAGT 1564
QY 616 ACCTGTGGAGACACAGGAGCTCTGGAGCCCTGGAGATGAATCAAGAGTCCCGGC 675
Db 1565 ATGTCAGAAGACTTTGACCGAGCCGTAGGCTGGTGAGCAGTGGACCCAGCGCTCCCCAC 1624
QY 676 GGCTGGAGAACTTCTCGAGAGACTTTGAGCTGCAGAGAGTGT---GTTACCTACCGCTCA 732
Db 1625 TGTTTAAAGACGTCRTCCACAGATCCAGAAGCAGGAGGTATCGGGAACCTGACGCTCC 1684
QY 733 ACACCTTCCTCTCGCGGCACTGCACCGGCTCATGCACTACAAAGCAGGTCTCGGAGCGGC 792
Db 1685 AGCACCACATGCTGGAGCCCGTGCAGAGGCTCCCGGTCAGGAGCTGCTCTCAAGGACT 1744
QY 793 TGTGCAAAACACCCCGCAGCGACCGCACTTCAGGAGACTGCCGAGCGGCTTTGSCAG 852
Db 1745 ATCTGAAGAGGCTCCCGAGAGCGCCCAAGACCGGAAGGATCGGAGAGGTCTTTGGAGC 1804
QY 853 AGATCAGGAGATGTTGGCAGAGCTCCAGGTACGATGATCAAGATGGAGAATTTCCAGA 912
Db 1805 TCATCTCCACAGCGCCCAACCACTCCAATGCTGCCATTCGGAAGTGGAGAAATGCACA 1864
QY 913 AGCTGCACGAACTCAAGAAGATTTGATTTGGCAATGACAACTTTGTTGTTCCGGGAAGG 972
Db 1865 AGCTCTTGGAGGTGTACGAGCAGCTGGTGGGAAGAAGACATTTGTCAACCGGCCAATG 1924
QY 973 AGTTCATCCGCTCGGCAGCCCTCAGCAAGCTCTCGGGGAG-----GGGCTCCACGAGC 1026
Db 1925 AACTGATCAAGGAGGGCCAAATCCAGAACTGTACGCAAGAACGCAACCCCGCAGGACC 1984
QY 1027 GCATGTTCTTCTTCAACGAGCTCTGCTATPACAGCGCGGGGCTGACGGCC---T 1083
Db 1985 GCCACCTCTCTCTTCAACAGCATGATCTTTACTGTGTGCCAAGCTCGCGCTCATGG 2044
QY 1084 CCAATCAGTTTAAAGTCCAGCGGCAGCTCCCGCTCTATGCAATGACGATGAGGAGAGCG 1143
Db 2045 GCCAGAAGTTTCAGGCTCCGGGGAAGATGGACATCTCAGGCTCCAGGCTCCAGGATATCG 2104

RESULT 13
US-09-799-799-1
; Sequence 1, Application US/09799799
; Patent No. US20020132291A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; FILE REFERENCE: CL001157
; CURRENT APPLICATION NUMBER: US/09/799,799
; CURRENT FILING DATE: 2001-03-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3380
; TYPE: DNA
; ORGANISM: Human
US-09-799-799-1

Query Match 5.6%; Score 64.4; DB 10; Length 3380;
Best Local Similarity 47.8%; Pred. No. 1.7e-08;
Matches 287; Conservative 0; Mismatches 301; Indels 12; Gaps 3;

QY 556 ACCAAAGATCGCGATGTCATCTGAAGNACATTCAGGCGATGAAGCACCTGGCGGCTC 615
Db 1225 ACCACGGCTCGGGACATCTTCAGAAGCTGGCCCATCTCTGAAAGATGTACGGCGAGT 1284
QY 616 ACCTGTGGAAGCACAGCGAGCGCTTGGAGCCCTGGAGATGAATCAAGAGCTCCCGGC 675
Db 1285 ATGTCAGAAGACTTTGACCGAGCCGTAGGCTGGTGAGCAGCTGGACCCAGCGCTCCCCAC 1344
QY 676 GGCTGGAGAACTTCTCGAGAGACTTTGAGCTGCAGAGAGTGT---GTTACCTACCGCTCA 732
```

```
Db 1345 TGTTTAAAGACGCTCGCTCCACAGCATCCAGAAGCAGGAGGTATCGGGAACTGACGCTGC 1404
QY 733 ACACCTTCTCTCTCGGGCCACTGCACCGGCTCATGCACTACAGCAGGTCCTGGAGCGGC 792
Db 1405 AGCACACATGCTGGAGCCCGTGCAGAGGTTCCCGGTTACGAGCTGCTGCTCAAGGACT 1464
QY 793 TGTGCAAAACACACCCCGCAGCGCACTTCAGGAGACTGCCGAGCGGCTTTGGGAC 852
Db 1465 ATCTGAAGAGGCTCCCGCAGGAGCGCCCAAGACCGGAAGGATCGGAGAGGTCTTTGGAGC 1524
QY 853 AGATCAGGAGATGTTGGCAGAGCTCCACGGTACGATGATCAAGATGGAGAATTTCCAGA 912
Db 1525 TCATCTCCACAGCGCCCAACCACTCCAATGCTGCCATTCGGAAGTGGAGAAATGCACA 1584
QY 913 AGCTGCACCAACTCAAGAAGATTTGATTTGGCAATTCACAACTCTGTGTTCCGGGAAGG 972
Db 1585 AGCTCTTGGAGGTGTACGAGCAGCTGGGTGGGAAGAAGACATTTGTCAACCGGCCAATG 1644
QY 973 AGTTCATCCGCTCGGGCAGCCCTCAGCAAGCTCTCGGGGAAG-----GGGCTCCACGAGC 1026
Db 1645 AACTGATCAAGGAGGGCCAAATCCAGAACTGTACGCAAGAACGCAACCCCGCAGGACC 1704
QY 1027 GCATGTTCTTCTTCAACGAGCTCTGCTATPACAGCGCGGGGCTGACGGCC---T 1083
Db 1705 GCCACCTCTCTCTTCAACAGCATGATCTTTACTGTGTGCCAAGCTCGCGCTCATGG 1764
QY 1084 CCAATCAGTTTAAAGTCCACGCGCAGCTCCCGCTCTATGCAATGAGGATGAGGAGAGCG 1143
Db 1765 GCCAGAAGTTTCAGCGTCCGGGAGAAAGATGGACATCTCAGGCTCCAGGTCAGGATATCG 1824

RESULT 14
US-09-764-868-316
; Sequence 316, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 316
; LENGTH: 1091
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1059)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-316

Query Match 5.6%; Score 64; DB 9; Length 1091;
Best Local Similarity 47.7%; Pred. No. 1.2e-08;
Matches 286; Conservative 1; Mismatches 301; Indels 12; Gaps 3;

QY 556 ACCAAAGATCGCGATGTCATCTGAAGAACAATTCAGGCGATGAAGCACCTGGCGGCTC 615
Db 188 ACCACGGCTCGGGACATCTTCAGAAGCTGGCCCATCTCTGAAAGATGTACGGCGART 247
QY 616 ACCTGTGGAAGCACAGCGAGCGCTTGGAGCCCTTGAGGCTGTGAGCAGCTGGAGCCAGCGCTCCCCAC 675
Db 248 ATGTCAGAAGACTTTGACCGAGCCGTAGGCTGGTGAGCAGCTGGAGCCAGCGCTCCCCAC 307
QY 676 GGCTGGAGAACTTCTCGAGAGACTTTGAGCTGCAGAGAGTGT---GTTACCTACCGCTCA 732
Db 308 TGTTTAAAGACGTCRTCCACAGCATCCAGAAGCAGGAGGTATCGGGAACCTGACGCTGC 367
QY 733 ACACCTTCTCTCTCGGGCCACTGCACCGGCTCATGCACTACAAAGCAGGTCTCGGAGCGGC 792
Db 368 AGCACCATGCTGTGGAGCCCGTGCAGAGGTTCCCGCGGTACGAGCTGCTGCTCAAGGACT 427
```

```
QY 793 TGTGCAAAACACCCCGCGAGCCAGCCGACTTCAGGGACTCCCGAGCGCTTTGGCAG 852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 ATCTGAAGAGGCTCCCGAGGAGCCCGACAGCCGGAAGGATCGGAGAGTCTCTTGSAGC 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 853 AGATCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGGAGAAATTCAGA 912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 TCATCTCCACAGCGCGCAACCACTCCCAATGCTGCCATTCGGAAGTGGAGAAATGCACA 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 913 AGCTGACGAACTCAAGAAAGATTTGATTGGCATTTGATGCTTGGTTCCCGGGAAGG 972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 AGCTTTGGAGGTGTAGGACAGCTGGTGGGAAGAAGACATTTGTAACCCCGCCAATG 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 973 AGTTTCATCCGTCGTGGGAGCGCTCAGCAAGCTCTCGGGGAAG-----GGGCTCCAGCAGC 1026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 AACTGATCAAGGAGGGCCAAATCCAGAAACTGTACGCAAGAACGGCACCCCCAGGACC 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1027 GCATGTTCTTCCTGTCAACAGAGTCTCTGTATACAGAGCCGGGGCTACGCGC---T 1083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 GCCACCTCTTCCTGTCAACAGCATGATCCTTTACTGTGTGCCCAAGCTCGGCTCATGG 727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1084 CCAATCAGTTTAAAGTCCAGCGGAGCTCCCGCTCTATGGCATGACGATTCAGGAGAGCG 1143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 GCCAAGATTCAGCGTCCGGGAGAAGATGCACATCTCAGGCTCCAGGTCAGGATATCG 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 15
US-09-815-242-7865
; Sequence 7865, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7865
; LENGTH: 3786
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(3786)
US-09-815-242-7865
```

Query Match 3.8%; Score 43.4; DB 10; Length 3786;  
Best Local Similarity 47.9%; Pred. No. 0.02;  
Matches 125; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
QY 623 GAAGCACAGCGAGCCCTTGGAGGCCCTTGGAGATCGAATGGAATCAAGAGCTCCCGCGGGTGG 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2541 GGAGCAGGACCTGCTACGGTGCCTGCTGATGACACCCCGAGCTGGCGGAGCC 2600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 683 GAACCTTCTGCAGAGACTTTGAGCTGCAGAGGTGTGTTACTTACCGCTCAACACCTTCCT 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2601 GTTCGGCGGCGAGCGGCATGACTGGAAGAAAGGCGAGTGCAGCCCGTCCCGGAGCGCAA 2660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 743 CCTCGCGGCCACTGCACCGGCTCATGCACTACAAGCAGGTCTCTGGAGCGGCTGTGCAACA 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2661 CCTCGCGGAGCTGCACCTGCTGAGCGCGACTACCCGAAAGCTTACCGCAAGTTACCTC 2720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 803 CCACCCCGGAGCGGAGCTTCAAGGACTCCCGAGCGCTTTGGCAGAGATCAGCGA 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2721 GCTCGGTCCGCTGCTGGACAAGCTGGGCAACGGCGGCAAGGGCATCGGCTGGAACACCG 2780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 863 GATGGTGGCACAGCTCCACGG 883
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2781 GAAGGAAGTGAAGCTGGTCGG 2801
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: December 3, 2002, 22:26:32  
Job time : 63.9989 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 16:31:10 ; Search time 1238.16 seconds  
(without alignments)  
15042.359 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_1352\_2501

Perfect score: 1150  
Sequence: 1 cgggtacaacgagcgac.....attgaggagcggaagacga 1150

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vrt:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	812.6	70.7	926	14	BQ706499
2	781.8	68.0	896	14	BQ672632
3	781	67.9	923	14	BQ950768
4	744.8	64.8	1043	14	BQ072025
5	726.2	63.1	901	12	BG475554
6	682	59.3	988	14	BQ674703

7	628.4	54.6	909	14	BQ646351
8	592	51.5	621	12	BG770181
9	591	51.4	861	12	BG764061
10	577.6	50.2	890	12	BE907778
11	574.8	50.0	784	12	BG763918
12	560.6	48.7	788	12	BG750463
13	551.2	47.9	902	12	BE910036
c	544	47.3	618	14	BQ331885
14	544	46.0	774	13	BI760610
15	529	43.4	932	12	BF686586
16	499	43.4	942	12	BM705217
17	494.4	43.0	496	14	BM705217
18	484.6	42.1	942	12	BG323704
19	475.4	41.3	489	9	AL121548
20	463	40.3	861	12	BG420356
21	457	39.7	1026	13	BM551766
22	454	39.5	454	14	BM728340
23	412.4	35.9	757	10	BE282962
24	398	34.6	719	12	BG248068
c	389.2	33.8	498	12	BF955437
26	368.2	32.0	703	13	BJ059899
27	366	31.8	449	10	BB840008
28	362.4	31.5	645	10	BE376581
29	356.6	31.0	861	12	BF538434
30	353	30.7	362	12	BF849713
31	350.2	30.5	565	13	BI663808
c	343.8	29.9	835	12	BE812222
33	340.6	29.6	838	12	BG829192
34	340.4	29.6	741	12	BG122769
35	339.4	29.5	436	12	BF955445
36	337.4	29.3	351	14	D81819
37	335.4	29.2	434	10	AW377949
38	331.6	28.8	616	12	BE912488
39	327.6	28.5	424	10	AW377971
40	326.8	28.4	466	14	R14694
41	323	28.1	516	13	BI898950
42	322.6	28.1	475	14	BQ301106
c	314.6	27.4	876	14	BQ221203
44	309.2	26.9	736	14	BM963756
45	304	26.4	414	10	BE271866

ALIGNMENTS

RESULT 1	BQ706499	926 bp	linear	EST 16-JUL-2002
LOCUS	AGENCOURT_8474935	NIH_MGC_113	Homo sapiens	cdna clone IMAGE:6301634
DEFINITION	5', mRNA sequence.			
ACCESSION	BQ706499			
VERSION	BQ706499.1	GI:21845398		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs@remail.nih.gov Tissue Procurement: Dr. Mark Watson CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Place: LNC2517 row: n column: 03 High quality sequence stop: 601. Location/Qualifiers			
FEATURES	1. .926			
source				

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6301634"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      242 a      247 c      254 g      183 t
ORIGIN
Query Match      70.7%; Score 812.6; DB 14; Length 926;
Best Local Similarity 98.8%; Pred. No. 2.7e-197;
Matches 829; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
QY 269 GACGACGATGAGATGAGCGCGGAGGAGAGATGCCCACTGATAAAGCGTACTTCAT 328
Db 1 GACGGACGATGAGGAGGCGCGGAGGAGAGATGCCCACTGATAAAGCGTACTTCAT 60
QY 329 AGCTAAGGAAGTGTCTACCAACGAGCGAATATCTGAAGGATCTCGAAGTTATCACTTC 388
Db 61 AGCTAAGGAAGTGTCTACCAACGAGCGAATATCTGAAGGATCTCGAAGTTATCACTTC 120
QY 389 GTGGTTTCAGAGCACAGTGAAGCAAGAGGACGCCATGCCGGAAGCACTGAAAGTCTCAT 448
Db 121 GTGGTTTCAGAGCACAGTGAAGCAAGAGGACGCCATGCCGGAAGCACTGAAAGTCTCAT 180
QY 449 ATTCCCGAATTTTGAACCTTGCACAAATTTTCATACATAATTTTCAAGAAATTCAGCA 508
Db 181 ATTCCCGAATTTTGAACCTTGCACAAATTTTCATACATAATTTTCAAGAAATTCAGCA 240
QY 509 ACGACTTGCCCTGTGGGAAGCGCGCTCAATGCCCAATACAGAGATTACCAAGAAATCGG 568
Db 241 ACGACTTGCCCTGTGGGAAGCGCGCTCAATGCCCAATACAGAGATTACCAAGAAATCGG 300
QY 569 CGATGTCTGCTGAAGAATTCAGGGCATGAAGCACTTGGCGGCTCACTGTGGAAGCA 628
Db 301 CGATGTCTGCTGAAGAATTCAGGGCATGAAGCACTTGGCGGCTCACTGTGGAAGCA 360
QY 629 CAGCAGGCGCTTGGAGCGCTCGAGATGGAATCAGAGCTCCCGCGCGTGGAGACTT 688
Db 361 CAGCAGGCGCTTGGAGCGCTCGAGATGGAATCAGAGCTCCCGCGCGTGGAGACTT 420
QY 689 CTGCAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCTCTCGG 748
Db 421 CTGCAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCTCTCGG 480
QY 749 GCACATGCACCGGCTCATGCACTACAGAGGTCTCGAGCGGCTGTGCAACACCAACCC 808
Db 481 GCACATGCACCGGCTCATGCACTACAGAGGTCTCGAGCGGCTGTGCAACACCAACCC 540
QY 809 GCCAGCCACGCGACTTCAGGAGCTCCGACCGCGCTTTGGCAGAGATCACGGAGATGT 868
Db 541 GCCAGCCACGCGACTTCAGGAGCTCCGACCGCGCTTTGGCAGAGATCACGGAGATGT 600
QY 869 GGCACAGCTCCACGATGATGATGAGAAATTTCCAGAACTGCACGAAGTCAA 928
Db 601 GGCACAGCTCCACGATGATGATGAGAAATTTCCAGAACTGCACGAAGTCAA 660
QY 929 GAAAGATTGATGGCATGACAAATCTTGGTTCGGGAAGGAGTTTCATCGTCTGGG 988
Db 661 GAAAGATTGATGGCATGACAAATCTTGGTTCGGGAAGGAGTTTCATCGTCTGGG 720
QY 989 CAGCCTCAGCAAGCTCTCGGGAAGGCGCTCCAGAGCGCATGTCTTCTGTTCAACCA 1048
Db 721 CAGCCTCAGCAAGCTCTCGGGAAGGCGCTCCAGAGCGCATGTCTTCTGTTCAACCA 780
QY 1049 CGTCTGTATACACGAG-CCGGGGGCTGACGGCCTCCCAATCAGTTTAAAGTCCACGGG 1106
```

```
Db 781 CGCTCGTCTATACAGAGCCGCGGGGTGACGCCCTCCAATCAGTTAAAGTCCACGGG 839
RESULT 2
LOCUS      BQ672632      896 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8185043 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6257346
5', mRNA sequence.
ACCESSION  BQ672632
VERSION    BQ672632.1 GI:21783466
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 896)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2411 row: h column: 19
            High quality sequence stop: 630.
FEATURES   Location/Qualifiers
            1..896
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:6257346"
             /clone_lib="NIH_MGC_102"
             /tissue_type="epidermoid carcinoma, cell line"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
             Site_2: EcoRI; cDNA made by oligo-dT priming.
             Directionally cloned into EcoRI/XhoI sites using the
             following 5' adaptor: GGCACGAG(G). Library constructed
             by Ling Hong in the laboratory of Gerald M. Rubin
             (University of California, Berkeley) using ZAP-cDNA
             synthesis kit (Stratagene) and Superscript II RT (Life
             Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 230 a 257 c 253 g 156 t
ORIGIN
Query Match      68.0%; Score 781.8; DB 14; Length 896;
Best Local Similarity 98.3%; Pred. No. 2.1e-189;
Matches 811; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
QY 15 GCGGACGAGCGCGCTCGGCGCCACGAGGAGGAGGAGTTCGTTAAGATAGGACC 74
Db 1 GCGGACGAGCGCGCTCGGCGCCACGAGGAGGAGGAGGAGTTCGTTAAGATAGGACC 60
QY 75 CAGCAGAGTAAACCTCAGCCCCCGCAGCGCAAGCACAGGCTCCCTGACTGGCAGTCTCTAC 134
Db 61 CAGCAGAGTAAACCTCAGCCCCCGCAGCGCAAGCACAGGCTCCCTGACTGGCAGTCTCTAC 120
QY 135 CTTTCGAGCTGTCTGTAACCTCGCAGGGGGGAGTGGCCCTGCCAACGTGACCTTGCT 194
Db 121 CTTTCGAGCTGTCTGTAACCTCGCAGGGGGGAGTGGCCCTGCCAACGTGACCTTGCT 180
QY 195 CCCAACCTCAGCCCCGACACCAAGCAGGCGCTCTCCCTTGATCAGCCCGCTGCTGAATGAC 254
Db 181 CCCAACCTCAGCCCCGACACCAAGCAGGCGCTCTCCCTTGATCAGCCCGCTGCTGAATGAC 240
QY 255 CAGGCGTCCCCCGGACGACGATGAGGATGAGGCGCGGAGGAGAGATTCACCACTGAT 314
Db 241 CAGGCGTCCCCCGGACGACGATGAGGATGAGGCGCGGAGGAGAGATTCACCACTGAT 300
```

```
QY 315 AAAGCGTACTTCATAGCTAAGGAAGTGTCTTACACCGAGGACATATCTGAAGGATCTC 374
|||||
Db 301 AAAGCGTACTTCATAGCTAAGGAAGTGTCTTACACCGAGGACATATCTGAAGGATCTC 360
QY 375 GAAGTTATCACTTCGTTGTTTCAGACACAGTGTAGCAAGAGAGACGCCATGCCGGAAGCA 434
|||||
Db 361 GAAGTTATCACTTCGTTGTTTCAGACACAGTGTAGCAAGAGAGACGCCATGCCGGAAGCA 420
QY 435 CTGAAAGTCTCATATATCCCGAATTTTGAACACCTTTGCACAAAATTTTCATATAATTTTCTC 494
|||||
Db 421 CTGAAAGTCTCATATATCCCGAATTTTGAACACCTTTGCACAAAATTTTCATATAATTTTCTC 480
QY 495 AAGGAATTTGAGCAAGACACTTGGCCCTGTGGGAAGCGCGTCAAAATGCCCAAAATCAGAGAT 554
Db 481 AAGGAATTTGAGCAAGACACTTGGCCCTGTGGGAAGCGCGTCAAAATGCCCAAAATCAGAGAT 540
QY 555 TACCAAGAAATCGGCATGTCTGCTGAAGAACATTTAGGGGCATGAAGCACCTGGCGGCT 614
Db 541 TACCAAGAAATCGGCATGTCTGCTGAAGAACATTTAGGGGCATGAAGCACCTGGCGGCT 600
QY 615 CACCTGTGGAAGCACAGCAGGCGCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGG 674
|||||
Db 601 CACCTGTGGAAGCACAGCAGGCGCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGG 660
QY 675 CGCCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAAC 734
|||||
Db 661 CGCCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAAC 720
QY 735 ACCTTCTCTCGGGCCACTGCACCGGCTTCATGCACTACAAGCAGAGTCTCTGG-AGCGGCT 793
Db 721 ACCTTCTCTCGGGCCACTGCACCGGCTTCATGCACTACAAGCAGGCGCTTGAAGCGGCT 780
QY 794 GTGCAAAACACACCGCCG-AGCCACGCGGACTTCAGGGAGCTGCC 837
Db 781 GTGCAAAACACCGCCGGAAGCGCGCAACITTCAGGGAACGCC 825

RESULT 3
LOCUS BQ950768
DEFINITION AGENCOURT_8842056 Lupski_sciatic_nerve Homo sapiens cDNA clone
IMAGE:6204242 5', mRNA sequence.
ACCESSION BQ950768
VERSION BQ950768.1 GI:22366246
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 923)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13625 row: d column: 03
High quality sequence stop: 585.
Location/Qualifiers
1..923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6204242"
/clone_lib="Lupski_sciatic_nerve"
/sex="male"
/tissue_type="sciatic nerve"
```

FEATURES  
source

```
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/Note="Vector: pCMW-SPORT6 (Life Technologies); Site_1:
Not1; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGCTCCG-3' and
5'-GACTAGTCTTAGATCCGAGCGGCCCTCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 238 a 257 c 239 g 189 t
ORIGIN
Query Match 67.9%; Score 781; DB 14; Length 923;
Best Local Similarity 96.0%; Pred. No. 3.4e-189;
Matches 834; Conservative 0; Mismatches 30; Indels 5; Gaps 3;
QY 97 CGCAGCCAAAGCAGAGCTCCCTGACTGGCAGTCTTCCAGCTTTCCGAGCTGTCTGTGAAC 156
Db 31 CTATATCCCAACAGAGCTCCCTGACTGGCAGTCTTCCAGCTTTCCGAGCTGTCTGTGAAC 90
QY 157 CGCAGGGGGAGTGGCCCTGCGCAACGTCGTCCTGCTGCCAACCTTGAGCCCGACACCA 216
Db 91 CGCAGGGGGAGTGGCCCTGCGCAACGTCGTCCTGCTGCCAACCTTGAGCCCGACACCA 150
QY 217 AGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGACAGGCGCTGCCCGGACGAGCG 276
Db 151 AGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGACAGGCGCTGCCCGGACGAGCG 210
QY 277 ATGAGATGAGGCGCGGAGGAGATTCCTCAACTGATAAAGCTACTTCATAGCTAAGG 336
Db 211 ATGAGATGAGGCGCGGAGGAGATTCCTCAACTGATAAAGCTACTTCATAGCTAAGG 270
QY 337 AAGTGTCTACCGGAGCAACATATCTGAAGGATCTGAAGTATTCATCTTCTGTGTTTC 396
Db 271 AAGTGTCTACCGGAGCAACATATCTGAAGGATCTGAAGTATTCATCTTCTGTGTTTC 330
QY 397 AGAGCACAGTGAGCAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATATCCCGA 456
Db 331 AGAGCACAGTGAGCAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATATCCCGA 390
QY 457 ATTTTGAACCTTTGACAAAATTTTCACTAATAATTTTCAAGGAAATTGAGCAACGACTTG 516
Db 391 ATTTTGAACCTTTGACAAAATTTTCACTAATAATTTTCAAGGAAATTGAGCAACGACTTG 450
QY 517 CCTGTGGGAAGCGCGCTCAATGCCCCAATCAGAGATTACCAAAAGATCGGCGATGTCA 576
Db 451 CCTGTGGGAAGCGCGCTCAATGCCCCAATCAGAGATTACCAAAAGATCGGCGATGTCA 510
QY 577 TGCTGAAGAACATTTACGGGCATGAAGCAGCTGGCGGCTCACCTGTGGAAGCAGCAGGAG 636
Db 511 TGCTGAAGAACATTTACGGGCATGAAGCAGCTGGCGGCTCACCTGTGGAAGCAGCAGGAG 570
QY 637 CCTTGGAGGCCCTTGGAGAATGGAATCAAGAGCTCCCGCGGCTGGAGAAGTCTTCGAGAG 696
Db 571 CCTTGGAGGCCCTTGGAGAATGGAATCAAGAGCTCCCGCGGCTGGAGAAGTCTTCGAGAG 630
QY 697 ACTTTGAGCTGCAGAAGTGTGTACTACCGCTCAACACCTTCTTCTTCTGCGGCCACTGC 756
Db 631 ACTTTGAGCTGCAGAAGTGTGTACTACCGCTCAACACCTTCTTCTTCTGCGGCCACTGC 690
QY 757 ACCGGCTCATGCACTCAAGCAGGCTCTTGAGGCGGCTGTGCAAAACACCAACCGCCCG-AG 814
Db 691 ACCGGCTCATGCACTCAAGCAGGCTCTTGAGGCGGCTGTGCAAAACACCAACCGCCCGAGC 750
QY 815 CCACGCCGACTTTCAGGAGCTGCCGAGCCCG-TTTGGCAGAGATCAC--GGAGATGTGTGC 871
Db 751 CACGCCGACTTTCAGGAGCTGCCGAGCCCGCTTTTGGCAGAGATCACCGGGAATGGGTGC 810
QY 872 ACAGTCCACGGTACGATGATCAAGATGAGAAATTTCCAGAGCTGCACGAACTCAAGAA 931
|||||
```

```

Db 811 ACAGCTCCACGGTACGATGAGATGAGAGAAATTTCCCGAAGCTGGCCCAACTCAGAAA 870
QY 932 AGATTTGATTGGCATTGACAATCTTGTGG 960
|||||
Db 871 AGATTTGATTGGCATTGACAATCTTGTGG 899
|||||

RESULT 4
BQ072025 1043 bp mRNA linear EST 02-APR-2002
LOCUS AGENCOURT_6859787 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5920511
DEFINITION 5', mRNA sequence.
ACCESSION BQ072025
VERSION BQ072025.1 GI:19901071
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1043)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2105 row: c column: 08
High quality sequence stop: 626.
Location/Qualifiers
1..1043
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5928511"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
Note="organ: brain; Vector: pORF7; Site_1: xhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 266 a 290 c 294 g 193 t
ORIGIN
Query Match 64.8%; Score 744.8; DB 14; Length 1043;
Best Local Similarity 99.7%; Pred. No. 7.1e-180;
Matches 746; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 390 TGGTTTCAGACACAGTGGACGAAGAGGAGCGATGCCGAGACACTGAAAAGTCTCAT 449
|||||
Db 1 TGGTTTCAGACACAGTGGACGAAGAGGAGCGATGCCGAGACACTGAAAAGTCTCAT 60
|||||
QY 450 TTCCCGAATTTTGACCTTTGCACAAATTCATACATACTAATTTCTCAAGAAATTTAGCAA 509
|||||
Db 61 TTCCCGAATTTTGACCTTTGCACAAATTCATACATACTAATTTCTCAAGAAATTTAGCAA 120
|||||
QY 510 CGACTTGCCTGTGGAGAGCGCGCTCAATGCCAAATCAGAGATTACCAAGAAATTCGCG 569
|||||
Db 121 CGACTTGCCTGTGGAGAGCGCGCTCAATGCCAAATCAGAGATTACCAAGAAATTCGCG 180
|||||
QY 570 GATGTCATCTGAAGACATTCAGGCGATGAAGACCTGGCGGCTCACCTGTGGAGCAC 629
|||||
Db 181 GATGTCATCTGAAGACATTCAGGCGATGAAGACCTGGCGGCTCACCTGTGGAGCAC 240
|||||

```

```

QY 630 AGCGAGGCTTTGGAGGCCCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 689
|||||
Db 241 AGCGAGGCTTTGGAGGCCCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 300
|||||
QY 690 TGCAGAGACTTTTGGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGGG 749
|||||
Db 301 TGCAGAGACTTTTGGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGGG 360
|||||
QY 750 CCACCTGCACCGGCTCATGCACCTACAAGCAGGTCTCTGGAGCGGCTGTGCAAAACACACCGG 809
|||||
Db 361 CCACCTGCACCGGCTCATGCACCTACAAGCAGGTCTCTGGAGCGGCTGTGCAAAACACACCGG 420
|||||
QY 810 CCGAGCCACCGCGCACTTCAGGGACTGCCAGCGCTTTGGCAGAGATCACGGAGATGGTG 869
|||||
Db 421 CCGAGCCACCGCGCACTTCAGGGACTGCCAGCGCTTTGGCAGAGATCACGGAGATGGTG 480
|||||
QY 870 GCACAGCTCCACGGTACGATGATCAAGATGAGAGATTTCCAGAAGCTGCACAACTCAAG 929
|||||
Db 481 GCACAGCTCCACGGTACGATGATCAAGATGAGAGATTTCCAGAAGCTGCACAACTCAAG 540
|||||
QY 930 AAAGATTTGATTGGCATTCACAATCTTGTGGTTCGGGAAGGAGTTCATCGTCTGGGC 989
|||||
Db 541 AAAGATTTGATTGGCATTCACAATCTTGTGGTTCGGGAAGGAGTTCATCGTCTGGGC 600
|||||
QY 990 AGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTCTGTTCAACGAC 1049
|||||
Db 601 AGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTCTGTTCAACGAC 660
|||||
QY 1050 GTCCTGCTATACACAGCGCGGGGCTGACGGGCTCCAAATCAGTTTAAAGTCACCGGGCAG 1109
|||||
Db 661 GTCCTGCTATACACAGCGCGGGGCTGACGGGCTCCCAATCAGTTTAAAGTCACCGGGCAG 720
|||||
QY 1110 CTCCTGCTCTATGGCATGACGATTGAGG 1137
|||||
Db 721 CTCCTGCTCTATGGCATGACGATTGAGG 748
|||||

RESULT 5
BQ475554 901 bp mRNA linear EST 21-MAR-2001
LOCUS 602491544F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4620346 5',
mRNA sequence.
ACCESSION BQ475554
VERSION BQ475554.1 GI:13407833
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 901)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTp
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1381 row: h column: 11
High quality sequence stop: 818.
Location/Qualifiers
1..901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4620346"
/clone_lib="NIH_MGC_20"
/lab_host="DH10B (phage-resistant)"
FEATURES
source

```

/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 232 a 243 c 253 g 173 t

ORIGIN

Query Match 53.1%; Score 726.2; DB 12; Length 901;  
Best Local Similarity 91.3%; Pred. No. 3.8e-175;  
Matches 817; Conservative 0; Mismatches 68; Indels 10; Gaps 4;

QY 229 CCTTGATCAGCCGCTGCTGAATGACCAGGCTGCCCGGAGCGGATGAGGATGAGG 288  
|||||  
Db 2 CCTTGATCAGCCGCTGCTGAATGACCAGGCTGCCCGGAGCGGATGAGGATGAGG 61  
|||||

QY 289 GCCGGAGGAAGAGATCCCAACTGATAAGCGTACTTCATAGCTAAGGAAGTGTCTACCA 348  
|||||  
Db 62 GCCGGAGGAAGAGATCCCAACTGATAAGCGTACTTCATAGCTAAGGAAGTGTCTACCA 121  
|||||

QY 349 CCAGGCGAATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCAGAGTGA 408  
|||||  
Db 122 CCAGGCGAATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCAGAGTGA 181  
|||||

QY 409 GCAAAGAGGAGCGCATGCCGGAAGCACTGAAAGTCTCATATTCCCGAATTTTGAACCTT 468  
|||||  
Db 182 GCAAAGAGGAGCGCATGCCGGAAGCACTGAAAGTCTCATATTCCCGAATTTTGAACCTT 241  
|||||

QY 469 TGCACAAATTTCTACTAATTTTCTCAAGGAATTTGAGCAACGACTTGCCTCTGGGAAG 528  
|||||  
Db 242 TGCACAAATTTCTACTAATTTTCTCAAGGAATTTGAGCAACGACTTGCCTCTGGGAAG 301  
|||||

QY 529 GCCGCTCAATGCCCAATCAGAGATTTACCAAGATTCGCGGATTCATGCTCAAGAAACA 588  
|||||  
Db 302 GCCGCTCAATGCCCAATCAGAGATTTACCAAGATTCGCGGATTCATGCTCAAGAAACA 361  
|||||

QY 589 TTCAGGCGATGAAGCAGCTGCCGCGCTCACCTGTGGAAGCACACGAGGCGCTTCGAGGCC 648  
|||||  
Db 362 TTCAGGCGATGAAGCAGCTGCCGCGCTCACCTGTGGAAGCACACGAGGCGCTTCGAGGCC 421  
|||||

QY 649 TGAGAAATGGAATCAAGAGCTCCGCGGCTGGAGAACTTCTCGAGAGCTTTGAGCTGC 708  
|||||  
Db 422 TGAGAAATGGAATCAAGAGCTCCGCGGCTGGAGAACTTCTCGAGAGCTTTGAGCTGC 481  
|||||

QY 709 AGAGGTGTGTACCTACCGCTCAACACCTTCTCTCTGGGGCCACTGCACCGGCTCATGC 768  
|||||  
Db 482 AGAGGTGTGTACCTACCGCTCAACACCTTCTCTCTGGGGCCACTGCACCGGCTCATGC 541  
|||||

QY 769 ACTACAAGCAGGTCTTGGAGCGGCTGTGCAAAACACACCGCGGAGCCAGCGGACTTCA 828  
|||||  
Db 542 ACTACAAGCAGGTCTTGGAGCGGCTGTGCAAAACACACCGCGGAGCCAGCGGACTTCA 601  
|||||

QY 829 -GGGACTGCCAGCCGCTTTGGCAGAGATCACGGAGATGGTGGCAGAGCTCCACGGTACG 887  
|||||  
Db 602 GGGCACTGGCAGCCGCTTTGGCAGAGATCACGGAGATGGTGGCAGAGCTCCACGGTACG 661  
|||||

QY 888 ATGATCAAGATGAGAAATTTCCAGAGCTGCACGAACTCAAGAAAGATTTGATTTGCATT 947  
|||||  
Db 662 ATGATCAAGATGAGAAATTTCCAGAGCTGCACGAACTCAAGAAAGATTTGATTTGCATT 721  
|||||

QY 948 GACAATCTGTGTGTTCCGGGAA---GGGAGTTTCATCCGCTCTGGGAGCGCTCAGCAA--- 1000  
|||||  
Db 722 GACAATCTGTGTGTTCCGGGAAAGCGGAGTTCCATCCGCTCTGGGAGCGCTTCAAGCAAGC 781  
|||||

QY 1001 -GCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTCTGTTCAACAGAGCTCCT-GCTA 1058  
|||||  
Db 782 TCCTCGGGGAAGGGACTCCACGAGCGCATGTTCTTCTCTGTTCAACAGAGCTCCTGGCTT 841  
|||||

QY 1059 TACACAGCGGGGGCTGACGGCCTCCAATACAGTTTAAAGTCACCGGGGAGCTCT 1113  
|||||

Db 842 TACACAGCGCGGGCTGACAGGCGTCCAATCCAGCTTAAGCCAAGGCGAGTCCC 896

RESULT 6  
BQ674703 988 bp mRNA linear EST 15-JUL-2002  
LOCUS AGENCOURT\_8189989 NIH\_MGC\_102 Homo sapiens cDNA clone IMAGE:6255217  
DEFINITION 5', mRNA sequence.

ACCESSION BQ674703  
VERSION BQ674703.1 GI:21785537  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 988)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LNCM2405 row: p column: 02  
High quality sequence stop: 685.

FEATURES  
Location/Qualifiers  
1..988  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6255217"  
/tissue\_type="epidermoid carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: salivary gland; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGCAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 216 a 306 c 291 g 161 t 14 others

ORIGIN

Query Match 59.3%; Score 682; DB 14; Length 988;  
Best Local Similarity 99.6%; Pred. No. 8.6e-164;  
Matches 693; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 456 AATTTTGAACCTTTGCACAAATTTTCATCTAATTTTCTCAAGAAATTTGAGCAACGACTT 515  
|||||  
Db 1 AATTTTGAACCTTTGCACAAATTTTCATCTAATTTTCTCAAGAAATTTGAGCAACGACTT 60  
|||||

QY 516 GCCTCTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAATCGCGGATGTC 575  
|||||

Db 61 GCCTCTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAATCGCGGATGTC 120  
|||||

QY 576 ATGCTGAAGAACAATTCAGGGCATGAAGCACCTGCGCGCTCACCTGTGGAAGCACACGCGAG 635  
|||||

Db 121 ATGCTGAAGAACATTCAGGGCATGAAGCACCTGCGCGCTCACCTGTGGAAGCACACGCGAG 180  
|||||

QY 636 GCCTTGGAGGCCCTCGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTTCGAGA 695  
|||||

Db 181 GCCTTGGAGGCCCTCGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTTCGAGA 240  
|||||

QY 696 GACTTTGAGCTGCAGAAAGGTGTTTACCTACCGCTCAACACCTTCTCTCTCGGGCCACTG 755  
|||||

Db 241 GACTTTGAGCTGCAGAAAGGTGTTTACCTACCGCTCAACACCTTCTCTCTCGGGCCACTG 300  
|||||

```
QY 756 CACCGCGCTCATGCACTACAAAGAGGTCTCGAGCGGCTGTGAAACACACACCGCGGAGC 815
Db 301 CACCGGCTCATGCACTACAAAGAGGTCTCGAGCGGCTGTGAAACACACACCGCGGAGC 360
QY 816 CAGCCGCACTTACGAGACTGCCAGCGGCTTTGGCAGAGATCACGAGATGTTGGCAGAG 875
Db 361 CAGCCGCACTTACGAGACTGCCAGCGGCTTTGGCAGAGATCACGAGATGTTGGCAGAG 420
QY 876 CTCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGCACGAATCTCAAGAAAGAT 935
Db 421 CTCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGCACGAATCTCAAGAAAGAT 480
QY 936 TTGATTGGCAATTGACAATCTTGTTGGTTCGGGAAGGAGTTTCATCGTCTGGGCAAGCTTC 995
Db 481 TTGATTGGCAATTGACAATCTTGTTGGTTCGGGAAGGAGTTTCATCGTCTGGGCAAGCTTC 540
QY 996 AGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTCCTG 1055
Db 541 AGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTCCTG 600
QY 1056 CTATACACGAG-CCGGGGGCTGACGGCTCCCAATCAAGTTTAAAGTCCACGGCAGCTCCC 1114
Db 601 CTATACACGAGNCCGGGGCTGACGGCTCCCAATCAAGTTTAAAGTCCACGGCAGCTCCC 560
QY 1115 GCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 1150
Db 661 GCTCTATGGCATGACGATTGANGAGAGCGAAGACNA 696

RESULT 7
BQ646351
LOCUS
DEFINITION BQ646351 909 bp mRNA linear EST 15-JUL-2002
AGENCOURT_8286354 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6299462
5', mRNA sequence.
ACCESSION BQ646351
VERSION BQ646351.1 GI:21770523
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 909)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCN2512 row: c column: 15
High quality sequence stop: 589.
FEATURES
Location/Qualifiers
1..909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6299462"
/clone_11b="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="pDL08 (phage-resistant)"
/note="organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
```

```
BASE COUNT 179 a 332 c 235 g 163 t
ORIGIN
Query Match 54.6%; Score 628.4; DB 14; Length 909;
Best Local Similarity 97.6%; Pred. No. 4.5e-150;
Matches 638; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 497 GGAATTGAGCAACGACTTGCCTGTGGGAAGCGGCTCAAATGCCAAATCAGAGATTGA 556
Db 1 GGAATTGAGCAACGACTTGCCTGTGGGAAGCGGCTCAAATGCCAAATCAGAGATTGA 60
QY 557 CCAAGAATTCGGCGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGCTCA 616
Db 61 CCAAGAATTCGGCGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGCTCA 120
QY 617 CCGTGTGGAGACACGCGAGCGCTTGGAGGCCCTGGAGAAATGAATCAAGAGCTCCGGCG 676
Db 121 CCGTGTGGAGACACGCGAGCGCTTGGAGGCCCTGGAGAAATGAATCAAGAGCTCCGGCG 180
QY 677 GCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACAC 736
Db 181 GCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACAC 240
QY 737 CTTCTCTCTGGCGGCACTGCACCGGCTCATGCACTACAAAGCAGGTCCTGGAGCGGCTGTG 796
Db 241 CTTCTCTCTGGCGGCACTGCACCGGCTCATGCACTACAAAGCAGGTCCTGGAGCGGCTGTG 300
QY 797 CAAACACACCGCGGAGCGGCTTCCAGGACTCCGAGCGGCTTCCAGAGAGAT 856
Db 301 CAAACACACCGCGGAGCGGCTTCCAGGACTCCGAGCGGCTTCCAGAGAGAT 360
QY 857 CACGAGATGTTGGCACACGCTCCAGGTCAGATGATCAAGATGGAGAAATTCAGAAGCT 916
Db 361 CACGAGATGTTGGCACACGCTCCAGGTCAGATGATCAAGATGGAGAAATTCAGAAGCT 420
QY 917 GCACGAACCTCAAGAAGATTTGATTGGCAATTCACAAATCTTGTGGTCCGGGAAGGAGTT 976
Db 421 GCACGAACCTCAAGAAGATTTGATTGGCAATTCACAAATCTTGTGGTCCGGGAAGGAGTT 480
QY 977 CATCGCTGCGGCGGCTCAGCAAGCTTCGGGGAAGGGGCTCCAGCAGCGATGTTCTTT 1036
Db 481 CATCGCTGCGGCGGCTCAGCAAGCTTCGGGGAAGGGGCTCCAGCAGCGATGTTCTTT 540
QY 1037 CCGTGTCAACGACGCTCCTGCTATACACGAGCGGGGCTGACGGCTCCCAATCAGTTTAA 1096
Db 541 CCGTGTCAACGACGCTCCTGCTATACACGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAA 600
QY 1097 AGTCCACGGGCGAGTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 1150
Db 601 AGTCCACGGGCGAGTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 654

RESULT 8
BQ770181
LOCUS
DEFINITION BQ770181 621 bp mRNA linear EST 15-MAY-2001
602744940F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4878177 5',
mRNA sequence.
ACCESSION BQ770181
VERSION BQ770181.1 GI:14080834
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 621)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCFD/PTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```





```

QY 919 ACGAACTCAAGAAAGATTGATTGGCATGTGACAACTTCTGTGTTCCTCGGGAAGGGAGTTCA 978
|||||
Db 362 ACGAACTCAAGAAAGATTGATTGGCATGTGACAACTTCTGTGTTCCTCGGGAAGGGAGTTCA 421
|||||
QY 979 TCCGTTCTGGGAGCCTCAGCAGACTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 1038
|||||
Db 422 TCCGTTCTGGGAGCCTCAGCAGACTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 481
|||||
QY 1039 TGTTCACACAGCTCTCTGTCTATACAGAGCGCGGGGCTGACGGCCTCCAATCAGTTTAAAG 1098
|||||
Db 482 TGTTCACACAGCTCTCTGTCTATACAGAGCGCGGGGCTGACGGCCTCCAATCAGTTTAAAG 541
|||||
QY 1099 TCCAGGGCAGCTCCCGCTCTATGTCATGACGATTGAGGAGCGAAGACGA 1150
|||||
Db 542 TCCAGGGCAGCTCCCGCTCTATGTCATGACGATTGAGGAGCGAAGACGA 593
|||||

RESULT 10
BE907778 890 bp mRNA linear EST 20-OCT-2000
LOCUS 601502061F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3903855 5',
DEFINITION mRNA sequence.
ACCESSION BE907778
VERSION BE907778.1 GI:10401682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 890)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9708 row: j column: 16
High quality sequence stop: 646.

FEATURES
Source
1..890
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3903855"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 251 a 242 c 241 g 156 t
ORIGIN
Query Match 50.2%; Score 577.6; DB 12; Length 890;
Best Local Similarity 98.4%; Pred. No. 4.7e-137;
Matches 657; Conservative 0; Mismatches 4; Indels 7; Gaps 7;

QY 485 TAATTTCTCAAGGAATGAGCAAGACTTGCCTGTGGGAGGCGCTCAATGCCCA 544
|||||
Db 1 TAATTTCTCAAGGAATGAGCAAGACTTGCCTGTGGGAGGCGCTCAATGCCCA 60
|||||
QY 545 AATCAGAGATTACCAAGAATCGGCATGTCATGCTGAAGAATTCAGGGCATGAAGCA 604
|||||
Db 61 AATCAGAGATTACCAAGAATCGGCATGTCATGCTGAAGAATTCAGGGCATGAAGCA 120
|||||
QY 605 CTGG-CCGCTCACCTGTGGGAAGCAGGAGGCTTGGAGGCCCTTGGAGATGAATCA 663
|||||

```

```

Db 121 CTGGACGGCTCACCTGTGGGAAGCACACGAGGGCTTGGAGGCGCTGGAGATGAATCA 180
QY 664 AGAGTCTCCGGCGGC-TGGAGAACTTC-TGCAGAGACTTTGAGCTGCAAGAGGTGTGTTA 721
|||||
Db 181 AGAGTCTCCGGCGGCATGGAGAACTTCATGACAGACTTTGAGCTGCAAGAGGTGTGTTA 240
|||||
QY 722 CTAACCG-CTCAACACCTTCTCTGCGGCCACTGCACCGGCTCATGCACTACAAGCAGG 780
|||||
Db 241 CCTACCGACTCAACACCTTCTCTGCGGCCACTGCACCGGCTCATGCACTACAAGCAGG 300
|||||
QY 781 TCCTGGAGCGGCTGTGCAAAACACACACCGCGGAGCCAGCGGACTTCAGGAGACTCCCGAG 840
|||||
Db 301 TCCTGGAGCGGCTGTGCAAAACACACACCGCGGAGCCAGCGGACTTCAGGAGACTCCCGAG 360
|||||
QY 841 CGCTTTTGGCAGAGATCACGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGG 900
|||||
Db 361 CGCTTTTGGCAGAGATCACGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGG 420
|||||
QY 901 AGAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGG 960
|||||
Db 421 AGAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGG 480
|||||
QY 961 TTCGGGAAGGAGTTTATCGTCTGGGCGAGCTCAGCAAGCTCTCGGGGAGGAGGCTCC 1020
|||||
Db 481 TTCGGGAAGGAGTTTATCGTCTGGGCGAGCTCAGCAAGCTCTCGGGGAGGAGGCTCC 540
|||||
QY 1021 AGCAGCGCATGTTCTCTCTGTTCAA-CGAGTCTCTGCTATACAGGCGCGGG-CTGAC 1078
|||||
Db 541 AGCAGCGCATGTTCTCTCTGTTCAA-CGAGTCTCTGCTATACAGGCGCGGGCTGAC 600
|||||
QY 1078 GGCCTTCCAAATCAGTTTAA-AGTCCACGGGCGAGCTCCCGCTCTATGGCATGACGATTGAGG 1137
|||||
Db 601 GGCCTTCCAAATCAGTTTAA-AGTCCACGGGCGAGCTCCCGCTCTATGGCATGACGATTGAGG 660
|||||
QY 1138 AGAGCGAA 1145
|||||
Db 661 AGAGCGAA 668

RESULT 11
BE9073918 784 bp mRNA linear EST 15-MAY-2001
LOCUS 602736888F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862444 5',
DEFINITION mRNA sequence.
ACCESSION BE9073918
VERSION BE9073918.1 GI:14074571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 784)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/FTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM1722 row: o column: 21
High quality sequence stop: 761.

FEATURES
source
1..784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4862444"
/clone_lib="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"

```



```

Db 481 CAGATCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAACTGCAGAACTCAGGAAA 540
QY 933 GATTTGATTGGCATTCACAACTCTGTGTTCCGG-GAAGGGAGTTTCATCCGCTCGGCGAG 991
Db 541 GATTGATTGGCATTCACAACTCTGTGTTCCGGGAGAGGGAGTTCAACCGTCTGGGCGAG 600
QY 992 CCTCAGCAAGCTCTCGGGGAA-GGGGCTCCAGCAG-CGCATGTTCTTCTGTTTCAACGAC 1049
Db 601 ACTCAGCAAGCTCTCGGGGAACGGGCTCCAGCAGCCGCGCATCTTCTGTTCAACGAC 660
QY 1050 GTCTGCTATACACAGCCGGGGCTGACGGGCTCAATCAGTTTAAAGTCACCGGGCAG 1109
Db 661 GTCTGCTATACACAGCCGGGGCTGACGG-CTCCAATCAGTTTAAAGTCACCGGGCAG 719
QY 1110 CTCCCCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 1150
Db 720 CTCCGATCTATGGCATGACGATTGAGGAGAGCGAAGACGA 760

RESULT 13
LOCUS BE910036 902 bp mRNA linear EST 20-OCT-2000
DEFINITION 601498085F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3900021 5',
mRNA sequence.
ACCESSION BE910036
VERSION BE910036.1 GI:10406227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9698 row: j column: 22
High quality sequence stop: 750.
FEATURES
source
Location/Qualifiers
1..902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3900021"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 231 a 248 c 272 g 151 t
ORIGIN
Query Match 47.9%; Score 551.2; DB 12; Length 902;
Best Local Similarity 94.9%; Pred. No. 2.8e-130;
Matches 592; Conservative 0; Mismatches 28; Indels 4; Gaps 2;

QY 1 CGGGTAACAGCAGCGGACGAGCGGCTCGCGCCCGCAGGAGGAGGAGGTCG 60
Db 161 CGGGTAACAGCAGCGGACGAGCGGCTCGCGCCCGCAGGAGGAGGAGGTCG 220
QY 61 TTAAGGATAGGACCCAGCAGAGTAACTCAGCCCCCGCAGCAGCAGGCTCCCTCA 120
Db 221 TTAAGGATAGGACCCAGCAGAGTAACTCAGCCCCCGCAGCAGCAGGCTCCCTCA 280

```

```

QY 121 CTGGCAGTCTCACCTTTCCGAGCTGTGTGTAACCTCGCAGGGGGAGTGGCCCCCTGCCA 180
Db 281 CTGGCAGTCTCACCTTTCCGAGCTGTGTGTAACCTCGCAGGGGGAGTGGCCCCCTGCCA 340
QY 181 ACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGAGCCCTCTCCCTTGTATCAGCC 240
Db 341 ACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGAGCCCTCTCCCTTGTATCAGCC 400
QY 241 CGCTGCTCAATGACACAGCGCTGCCCGCGAGCGAGCATGAGGATGAGGCGCGGAGGAAGA 300
Db 401 CGCTGCTCAATGACACAGCGCTGCCCGCGAGCGAGCATGAGGATGAGGCGCGGAGGAAGA 460
QY 301 GATTCCCAACTGATAAAGCTACTTTCATAGCTAAAGAAAGTGTCTACCAACCGAGCGAAT 360
Db 461 GATTCCCAACTGATAAAGCTACTTTCATAGCTAAAGAAAGTGTCTACCAACCGAGCGAAT 520
QY 361 ATCTGAAGGATCTCGAAGTTTATCACTTCGTGGTTTCAGAGCACAGTGAGCAAGAGGACG 420
Db 521 ATCTGAAGGATCTCGAAGTTTATCACTTCGTGGTTTCAGAGCACAGTGAGCAAGAGGACG 580
QY 421 CCATGCCGG-AGCACTGAAA--GTCTCATATATCCCGAATTTTGAACCTTTCACAAA 476
Db 581 CCATGCCGGAAAGCACTGAAAAGTCTCATATTTCCCGAATTTTGAACCTTTCACAAA 640
QY 477 TTTTACTATAATTTTCTCAAGGAATTTGAGCAACGACTTGGCCTGTGGGAAGGCCGCTCA 536
Db 641 TTTTACTATAATTTTCTCAAGGAATTTGAGCAACGACTTGGCCTGTGGGAAGGCCGCTCA 700
QY 537 AATGCCCAAACTCAGAGATTACCAAGAAATCGCGCATGTCTGCTCAAGAAATTCAGGGC 596
Db 701 AATGCCCAAACTCAGAGATTACCAAGAAATCGCGCATGTCTGCTCAAGAAATTCAGGGC 760
QY 597 ATGAGCAGCTTGGCGCTCACCTG 620
Db 761 ATAAAGCACCGGGGGTCCCTGTG 784

RESULT 14
LOCUS BQ331885/6 618 bp mRNA linear EST 17-MAY-2002
DEFINITION PM3-ET0268-100501-005-d11 ET0268 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ331885
VERSION BQ331885.1 GI:20973153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 618)
AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 2002663
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3st2=PM3-ET0268-
100501-005-d11&t3=2001-05-10&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 572.

```

FEATURES	Location/Qualifiers		REFERENCE	
	source		1 (bases 1 to 774)	
FEATURES	1..618		NIH-MGC http://mgc.nci.nih.gov/.	
	/organism="Homo sapiens"		National Institutes of Health, Mammalian Gene Collection (MGC)	
FEATURES	/db_xref="taxon:9606"		Unpublished (1999)	
	/clone_lib="ET0268"		Contact: Robert Strausberg, Ph.D.	
FEATURES	/dev_stage="Adult"		Email: c9apbs-r@mail.nih.gov	
	/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		Tissue Procurement: Life Technologies, Inc.	
FEATURES	128 a 162 c 159 g 169 t		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
	BASE COUNT		DNA Sequencing by: Incyte Genomics, Inc.	
FEATURES	128 a 162 c 159 g 169 t		Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
	BASE COUNT		High quality sequence stop: 774.	
FEATURES	128 a 162 c 159 g 169 t		Location/Qualifiers	
	ORIGIN		1..774	
FEATURES	47.3%; Score 544; DB 14; Length 618;		/organism="Homo sapiens"	
	Query Match		/db_xref="taxon:9606"	
FEATURES	98.6%; Pred. No. 1.5e-128;		/clone_lib="NIH_MGC_116"	
	Best Local Similarity		/lab_host="DH10B"	
FEATURES	580; Conservative 0; Mismatches 5; Indels 3; Gaps 3;		/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH MGC Library."	
	Matches 580; Conservative 0; Mismatches 5; Indels 3; Gaps 3;		BASE COUNT	
FEATURES	350 CGAGCGAATATCTGAAGGATCTCGAAGTATATCACTTCGTGGTTTCAGAGCACAGTGAG 409		200 a 214 c 226 g 134 t	
	Db		ORIGIN	
FEATURES	616 CGTGGCTACATATCGAAGGATCTCGAAGTATATCACTTCGTGGTTTCAGAGCACAGTGAG 558		Query Match	
	Qy		46.0%; Score 529; DB 13; Length 774;	
FEATURES	410 CAAAGAGGAGCGGATCGCGGAGCACTGAAAAGTCTCATATTCCCGAAATTTTGAACCTTT 469		Best Local Similarity	
	Db		100.0%; Pred. No. 1.2e-124;	
FEATURES	557 CAAAGAGGAGCGGATCGCGGAGCACTGAAAAGTCTCATATTCCCGAAATTTTGAACCTTT 499		Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
	Qy		Qy	
FEATURES	470 GCACAAATTTCTACTAATTTTCTCAAGGAATTTGAGCAACGACTTGCCTGTGGGAAGG 529		Db	
	Db		Qy	
FEATURES	498 GCACAAATGTCATCTAATTTTCTCAAGGAATTTGAGCAACGACTTGCCTGTGGGAAGG 439		Db	
	Qy		Qy	
FEATURES	530 CGCTCAAAATGCCAAATCAGAGATTACCAAGAATCGCGGATGTCATGCTGAAGAACAT 589		Db	
	Db		Qy	
FEATURES	438 CCCTCAATGCCAAATCAGAGATTACCAAGAATCGCGGATGTCATGCTGAAGAACAT 379		Db	
	Qy		Qy	
FEATURES	590 TCAGGCGATGAAGCACCTGGCGGCTACCTGTGGGAAGCACAGCGGCTTGGAGGCCCT 649		Db	
	Db		Qy	
FEATURES	378 TCAGGCGATGAAGCACCTGGCGGCTACCTGTGGGAAGCACAGCGGCTTGGAGGCCCT 319		Db	
	Qy		Qy	
FEATURES	650 GGAGAAATGAATCAGAGCTCCCGGGCTGGAGACTTCTCGAGAGACTTTGAGCTGCA 709		Db	
	Db		Qy	
FEATURES	318 GGAGAAATGAATCAGAGCTCCCGGGCTGGAGACTTCTCGAGAGACTTTGAGCTGCA 259		Db	
	Qy		Qy	
FEATURES	710 GAAGTGTGTTTACCTACCGCTCAACACCTTCTCTGCGGCCACTGCACCGGCTCATGCA 769		Db	
	Db		Qy	
FEATURES	258 GAAGTGTGTTTACCTACCGCTCAACACCTTCTCTGCGGCCACTGCACCGGCTCATGCA 199		Db	
	Qy		Qy	
FEATURES	770 CTACAAGCAGTCTGGAGCGGCTGTGCAACACCAACCGCGAGCCAGCGACTTCAG 829		Db	
	Db		Qy	
FEATURES	198 CTACAAGCAGTCTGGAGCGGCTGTGCAACACCAACCGCGAGCCAGCGACTTCAG 139		Db	
	Qy		Qy	
FEATURES	830 GGACTGCCAGCGGCTTTGGCAGAGATCAGGAGATGGTGGCAGACGCTCCAGGTTACGAT 889		Db	
	Db		Qy	
FEATURES	138 GGACTGCCAGCGGCTTTGGCAGAGATCAGGAGATGGTGGCAGACGCTCCAGGTTACGAT 79		Db	
	Qy		Qy	
FEATURES	890 GATCAAGTGGAGAAATTTCCAGAA-GCTGCAGCAACTCAAGAAAGATT 936		Db	
	Db		Qy	
FEATURES	78 GATCAAGTGGAGAAATTTCCAGAAATGCTCCAGAACTCAAGAAAGATT 31		Db	
	Qy		Qy	
FEATURES	774 bp mRNA linear EST 25-SEP-2001		Db	
	LOCUS		Qy	
FEATURES	BI760610		Db	
	DEFINITION		Qy	
FEATURES	603044980F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5185411 5', mRNA sequence.		Db	
	ACCESSION		Qy	
FEATURES	BI760610		Db	
	VERSION		Qy	
FEATURES	BI760610.1 GI:15752188		Db	
	KEYWORDS		Qy	
FEATURES	EST.		Db	
	SOURCE		Qy	
FEATURES	human.		Db	
	ORGANISM		Qy	
FEATURES	Homo sapiens		Db	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		Qy	
FEATURES	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		Db	
	ATACTAATTTTCTCAAGGAATTTGAGCAACGACTTGCCTGTGGGAAGG 529		Qy	

Db 726 ATACTAATTTCTCAAGGAAATTGAGCAACGACTTCCCTGTGGGAAGG 774  
|||||

Search completed: December 3, 2002, 21:25:20  
Job time : 1257.16 secs



protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors  
Biochem Biophys. Res. Commun. 241 (2), 369-375 (1997)  
98086358  
2 (bases 1 to 3442)  
Koyano,Y., Kawamoto,T. and Kato,Y.  
Direct Submission  
Submitted (22-Oct-1997) Takeshi Kawamoto, Hiroshima University School of Dentistry, Department of Biochemistry; 1-2-3 Kasumi Minami-ku, Hiroshima, Hiroshima 734, Japan  
(E-mail:tkawamo@ipc.hiroshima-u.ac.jp, Tel:082-257-5688, Fax:082-257-5629)

FEATURES

source  
1. .3442  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="chondrocyte"  
/tissue\_type="cartilage"  
/dev\_stage="embryo"  
49. .3186  
/function="Rho Guanine Nucleotide Exchange Factor"  
/note="Band 4.1 superfamily"  
/codon\_start=1  
/product="CDEP"  
/protein\_id="BAA24267.1"  
/db\_xref="GI:2766165"

CDS

QMLDDTQFAFVQAPKQVLLDVAQNLHNLVEGDIYFGLFFPDHKKRLTWNLDLLAPIV  
KQIRPKHVVVFVKFFPDHDLQBELRYFALQVKQDLAQGLKLTNCDTSAALLI  
SHVQSEIGDFEALDREHLAKNKYIQPDAALEDKIVFEFHNNHIGQTPAESDFLLEI  
ARLMYGIIRLPKADREGTKINLAVANTGILVFOGFKINAFNWKVPKLSFKRKF  
LTKLPDANSAYODTLEFLMASRDFCKSEWKTICVBEHAFRLFEPEPKPKPVLFSRG  
SSFRSGTKOVLVDYVKEGCHKVQFEPKHKSHIHSIRLSAQTELSNVELEQSQS  
TSLTGBEAESGSGSCRGKPKVSEPEGHPSPAPRSPAGNKQADGAAASATEE  
EEVVLTDQKQPPQSTGLTSGPSHLSVNSQGVAPANTVSPNLSPDTKQ  
ASPLSLNDQACPTDDEGRKRFPDRAKYFAKEVSTERTYI LKDEVIITSWF  
QSTVKEADMPALAKLIFPNPEPLKHFNFNKEIQRLALWEGRSNAQIRDYQRIQ  
DVMKNIQCMHAAHLKWHSEALEALEIKSSRRLENFCRDFLOKVCYUPLNTFL  
LRPLRLHYKQVLRCLKHPPSHADPRDCAALAEITEMVAQLHGTMIKWNFKL  
HELKDLIDGLNVYGRFIRLGLSLKSLGKGLQORMFFLENDVLLTTSRLTASNQ  
FKVHQPLDYGIMTIESEDEWGPCHLLILRGQSLIIVAASSRSSEMKWVEDIQMAD  
LAEKSSPAPELASPDNKPDEATAQDESDLLASRTSLERQAPHRGNTMVHV  
CWHRTSYSMVDFSTAVENQLSGNLLRKFKNNGWQKLWVFTNCLFFYKSHQDNHP  
LASPLGLYSILIPSESENIQDYVFKLHFKSHVYFRAESEYTFERWNEVIRSATS  
ASRPHVLSHKSILVY"  
3442  
polyA\_site  
/note="50 a nucleotides"  
BASE COUNT 864 a 952 c 927 g 699 t  
ORIGIN

Query Match 100.0%; Score 2351; DB 9; Length 3442;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTCAGAAAACCTGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60  
Dbl 151 CTTTCAGAAAACCTGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 210  
QY 61 GAAGTTCCCAAGAGCTCTGGGAGGTGCTGCTGGATGCAGTTTGCACCAACCTCAAC 120  
Dbl 211 GAAGTTCCCAAGAGCTCTGGGAGGTGCTGCTGGATGCAGTTTGCACCAACCTCAAC 270  
QY 121 CTGCTGGAAGGTGACTATTTGGCCCTCGAGTTTCCTGATCACAAAAAGATCACGGTGTGG 180  
Dbl 271 CTGCTGGAAGGTGACTATTTGGCCCTCGAGTTTCCTGATCACAAAAAGATCACGGTGTGG 330  
QY 181 CTGGATCTCCTAAACCCATTGTGAAACAGATTAGAAGGCCAAAGCACGCTGTGTTAAAG 240  
Dbl 331 CTGGATCTCCTAAACCCATTGTGAAACAGATTAGAAGGCCAAAGCACGCTGTGTTAAAG 390  
QY 241 TTTGTGGTGAATTCCTTCGGCCTGACCACACACACTCCAGAAGAACTCACAAGGTAC 300  
|||||

Dbl 391 TTTGTGGTGAATTCCTTCGGCCTGACCACACACAACTCCAAGAAGAACTCACAAGGTAC 450  
QY 301 CTGTTCCGCGCTGCAGGTGAAGCAGGACTTGGCTCAAGGAGGTTGACGTGTAATGACACC 360  
|||||  
Dbl 451 CTGTTCCGCGCTGCAGGTGAAGCAGGACTTGGCTCAAGGAGGTTGACGTGTAATGACACC 510  
QY 361 AGCGCAGCTCTCTTGTGATTTTACACACATTGTCGAATCTGAGATTGGGGATTTTGTATGAAGCC 420  
|||||  
Dbl 511 AGCGCAGCTCTCTTGTGATTTTACACACATTGTCGAATCTGAGATTGGGGATTTTGTATGAAGCC 570  
QY 421 TTGGCAGAGAGCAGCTTAGCAAAAAATAATACATACCTCAGCAAGCAGCAGTAGAGGAC 480  
|||||  
Dbl 571 TTGGCAGAGAGCAGCTTAGCAAAAAATAATACATACCTCAGCAAGCAGCAGTAGAGGAC 630  
QY 481 AAAATCGTGAATTTTACCATTACCATGGAACAACACACAGCAAACTCAGATTTCCAG 540  
|||||  
Dbl 631 AAAATCGTGAATTTTACCATTACCATGGAACAACACACAGCAAACTCAGATTTCCAG 690  
QY 541 CTCCTAGAGATTGCCCCGCTGAGATGATGGAATCGGTTGACCCGCGGCAAGGAC 600  
|||||  
Dbl 691 CTCCTAGAGATTGCCCCGCTGAGATGATGGAATCGGTTGACCCGCGGCAAGGAC 750  
QY 601 AGGAAGGCACGAAGATCAATCTGCGCTGAGATGATGGAATCGGTTGACCCGCGGCAAGGAC 660  
|||||  
Dbl 751 AGGAAGGCACGAAGATCAATCTGCGCTGAGATGATGGAATCGGTTGACCCGCGGCAAGGAC 810  
QY 661 TTCACCTAAGATCAATCGCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGAAG 720  
|||||  
Dbl 811 TTCACCTAAGATCAATCGCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGAAG 870  
QY 721 CGCTTTCTCATCAAGCTCGCGGCAGATGCCAATAGTCGCTACCAAGATACCTTTGGAATTC 780  
Dbl 871 CGCTTTCTCATCAAGCTCGCGGCAGATGCCAATAGTCGCTACCAAGATACCTTTGGAATTC 930  
QY 781 CTGATGGCAGTCGGGATTTCTGCAAGTCCTTCTCGAAAATCTGTTTGAACATCATGCC 840  
Dbl 931 CTGATGGCAGTCGGGATTTCTGCAAGTCCTTCTCGAAAATCTGTTTGAACATCATGCC 990  
QY 841 TTTCTTTAGACTTTTTCAGAGCCCAACCAAGCCCAAGCCCTCTTTAGCCGGGG 900  
Dbl 991 TTTCTTTAGACTTTTTCAGAGCCCAACCAAGCCCAAGCCCTCTTTAGCCGGGG 1050  
QY 901 TCATCATTTTCGGTTTTCAGTTCGCGACTCAGAACAGGATTCGCTGACTATGTTAAAGAAGA 960  
Dbl 1051 TCATCATTTTCGGTTTTCAGTTCGCGACTCAGAACAGGATTCGCTGACTATGTTAAAGAAGA 1110  
QY 961 GGACATAAGAGGTGTCAGTTTGAAGGAAGCAGACAGCAAGATTCATTTATCCGGAGCCCTT 1020  
Dbl 1111 GGACATAAGAGGTGTCAGTTTGAAGGAAGCAGACAGCAAGATTCATTTATCCGGAGCCCTT 1170  
QY 1021 GCTTCACGCCTACAGAACTGAATTCGGAAGTCTCGAGCAGTCTCAGCAGACAGCAGCAGC 1080  
Dbl 1171 GCTTCACGCCTACAGAACTGAATTCGGAAGTCTCGAGCAGTCTCAGCAGACAGCAGCAGC 1230  
QY 1081 CTTACATTTGGAGAAGGTCCGGAATCTCAGGGGGCCAGAGCTGCGCGGAGGAGGAGGAGGAA 1140  
Dbl 1231 CTTACATTTGGAGAAGGTCCGGAATCTCAGGGGGCCAGAGCTGCGCGGAGGAGGAGGAGGAA 1290  
QY 1141 CCGAAGTTTCCCGCGGAGCCGGGTGCGACCCAGCCCTGCGCCGAGGAGGAGGAGGAGGAGGAG 1200  
Dbl 1291 CCGAAGTTTCCCGCGGAGCCGGGTGCGACCCAGCCCTGCGCCGAGGAGGAGGAGGAGGAGGAG 1350  
QY 1201 GCGGGTAAACAAGCAGCGGAGCGGCGCTCGGCGCCACGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260  
Dbl 1351 GCGGGTAAACAAGCAGCGGAGCGGCGGCTCGGCGCCACGAGGAGGAGGAGGAGGAGGAGGAGGAG 1410  
QY 1261 GTTAAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCCGCAGCCAAAGCAGCAGAGCTCCCTG 1320  
Dbl 1411 GTTAAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCCGCAGCCAAAGCAGCAGAGCTCCCTG 1470  
QY 1321 ACTGGCAGTCTTCACCTTTCCGAGCTGCTGTGAACCTCGAGGGGGAGTGGCCCTGCC 1380  
Dbl 1471 ACTGGCAGTCTTCACCTTTCCGAGCTGCTGTGAACCTCGAGGGGGAGTGGCCCTGCC 1530  
|||||

QY	1381	AACTGACCTTGCTCCCAACCTGAGCCCGACACCAAGCAGGCCTCTCCCTTGATCAGC	1440
Db	1531	AACTGACCTTGCTCCCAACCTGAGCCCGACACCAAGCAGGCCTCTCCCTTGATCAGC	1590
QY	1441	CCGCTGCTGAATGACCAAGCCTGCCCGGACCGACGATGAGGCGCGGAGGAG	1500
Db	1591	CCGCTGCTGAATGACCAAGCCTGCCCGGACCGACGATGAGGCGCGGAGGAG	1650
QY	1501	AGATTCCCACTGATAAGCGTACTTCATAGCTAAGGAAGTCTACACCGAGCGGAACA	1560
Db	1651	AGATTCCCACTGATAAGCGTACTTCATAGCTAAGGAAGTCTACACCGAGCGGAACA	1710
QY	1561	TATCTGAAGATCTGAAGTTATCACTTCGTGTTTCAGAGCACAGTGAGCAAGAGGAC	1620
Db	1711	TATCTGAAGATCTGAAGTTATCACTTCGTGTTTCAGAGCACAGTGAGCAAGAGGAC	1770
QY	1621	GCCATGCCGAAGCACTGAAAGCTTCATATCCCGAATTTTGNACCTTTGCACAAATTT	1680
Db	1771	GCCATGCCGAAGCACTGAAAGCTTCATATCCCGAATTTTGNACCTTTGCACAAATTT	1830
QY	1681	CATACTAATTTTCTCAAGGAAATTTAGCAACGACTTGCCCTGTGGGAAGCCGCTCAAT	1740
Db	1831	CATACTAATTTTCTCAAGGAAATTTAGCAACGACTTGCCCTGTGGGAAGCCGCTCAAT	1890
QY	1741	GCCCAATCAGAGATTACCAAGAAATCGGCGATGTCTATGCTGAAGAACATTCAGGCGATG	1800
Db	1891	GCCCAATCAGAGATTACCAAGAAATCGGCGATGTCTATGCTGAAGAACATTCAGGCGATG	1950
QY	1801	AAGCACTGCGGCTCACCTGTGGAAGCACAGGAGGCCCTTGAGGCCCTGGAGATGGA	1860
Db	1951	AAGCACTGCGGCTCACCTGTGGAAGCACAGGAGGCCCTTGAGGCCCTGGAGATGGA	2010
QY	1861	ATCAAGAGCTCCCGGGGCTGGAGAACTTCTGCAGAGACTTTTCAGCTGCAGAAAGTGTGT	1920
Db	2011	ATCAAGAGCTCCCGGGGCTGGAGAACTTCTGCAGAGACTTTTCAGCTGCAGAAAGTGTGT	2070
QY	1921	TACCTACCGCTCAACACTTCTCTCGGGCCACTGACCGGCTCATGCACTACAGCAG	1980
Db	2071	TACCTACCGCTCAACACTTCTCTCGGGCCACTGACCGGCTCATGCACTACAGCAG	2130
QY	1981	GTCTTGAGCGGCTGTGCAAAACACCCCGCCGACGACCCGCTTCAGGGACTGCCGA	2040
Db	2131	GTCTTGAGCGGCTGTGCAAAACACCCCGCCGACGACCCGCTTCAGGGACTGCCGA	2190
QY	2041	GCGCTTTGGCAGAGATCAGGAGATGGTGGCAGCTCCACCGTACGATGATCAAGATG	2100
Db	2191	GCGCTTTGGCAGAGATCAGGAGATGGTGGCAGCTCCACCGTACGATGATCAAGATG	2250
QY	2101	GAGAAATTCAGAGCTGCACGAACTCAAGAAAGATTTGATTGGCATTTGCAATCTTGTG	2160
Db	2251	GAGAAATTCAGAGCTGCACGAACTCAAGAAAGATTTGATTGGCATTTGCAATCTTGTG	2310
QY	2161	GTTCGGGAGGAGGATTCATCCCTCTGGCGAGCCTCAGCAAGCTTCGGGGAGGGGCTC	2220
Db	2311	GTTCGGGAGGAGGATTCATCCCTCTGGCGAGCCTCAGCAAGCTTCGGGGAGGGGCTC	2370
QY	2221	CAGCAGCGATGTTCTCTCTGTTCAACAGCTGCTGCTATACACGAGCGGGGCTGACG	2280
Db	2371	CAGCAGCGATGTTCTCTCTGTTCAACAGCTGCTGCTATACACGAGCGGGGCTGACG	2430
QY	2281	GCTTCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAG	2340
Db	2431	GCTTCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAG	2490
QY	2341	AGCGAAGACCA 2351	
Db	2491	AGCGAAGACCA 2501	
RESULT 2	BC030329		
LOCUS		3902 bp	mRNA linear ROD 07-AUG-2002

DEFINITION

Mus musculus, clone IMAGE:5376197, mRNA, partial cds.

ACCESSION

BC030329.1

VERSION

GI:20987935

KEYWORDS

house mouse.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Strausberg, R.

REFERENCE

1 (bases 1 to 3902)

AUTHORS

Direct Submission

TITLE

Submitted (07-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

JOURNAL

NIH-MGC Project URL: http://mgc.nci.nih.gov

REMARK

Contact: MGC help desk

COMMENT

Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: http://www.nisc.nih.gov/  
Contact: nisc.mgc@nhgri.nih.gov  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgueon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

FEATURES

source

1..3902

/organism="Mus musculus"

/db\_xref="taxon:10090"

/map="FVB/N-3"

/clone="IMAGE:5376197"

/tissue\_type="Mammary tumor, MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."

/clone\_lib="NCI\_CGAP\_Mam2"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

<1..2510

/codon\_start=3

/product="Unknown (protein for IMAGE:5376197)"

/protein\_id="AAH30329.1"

/db\_xref="GI:20987936"

/translation="TRPEVARLEMYGIRLHPAKDREGTKINLAVANTGILVFGFTKINAFNWKVKLSFKKRLIKLRDPVNSYQDTLEFLMAGDFCKSPFKICVHHAPFLFEPPKPKPVLFSGSSPFRGTONQVLDIVKGGHKVQFERKHSKIHSTRLSVSPAPNSKVPKPSQASLTFGESTPGGSCQQAQAKETKACTLELPHQSPALPKSPGSKADGTTVPVPEEEEGKQDGRPSNPQPPQSTGSLTSPHLSELINSQCGAANVTLSPLNDQACPRTDDEEGRKRPTDKAYIAKEVSTERTYLDLEVIASFQSTVSKDSMEPEALSLIFPNFELPHFTFLAKEIEQRLALNRSNAHVQDQVRIQDMVKNIQGMKHLAHLWKHSEALEALELSIKSRRLEHFCRDFELQVYLPNTFLRLPHRLMHYKHLERLCKHPNHDPRDRAALAETEMVAOLHGTMIKKNFQKHLKDLIGDNLVTGPGREFIRLGSLSKSGLOQRMFFLNDVLLYTSRGLTASNQFKVHGQPLPYMTIESEEWGPHCITLRGQRQSIIVASSRSEMEKWMEDIOAIDLAKSNGPTPELLASSPDKNSPDNKAADQSEDDLSASRTSLERQAPHRGNTMVHVHRSSTVSMVDFSIAVENOLSGNLLKFKNSNGWKLWVYTFNCLFFYKSHQSDHPLSLPLGYSLTIPSESENHKKDYVFKLHFKSHVYYFRASEYTFERWMEVIRSATSSASRAHILSHKESHLY"

BASE COUNT

982 a 1083 c 997 g 840 t



ORIGIN

Query Match 57.5%; Score 1351; DB 10; Length 3902;  
Best Local Similarity 84.7%; Pred. No. 0;  
Matches 1542; Conservative 0; Mismatches 270; Indels 9; Gaps 2;

QY 540 GTCTCTAGAGATTGCCCGCTGGCTAGAGATGTATGGAATCCGGTTGCACCCGGCCAAAGGA 599  
DB 5 GCGTCGGAGGTGGCCGCTGCGCTAGAGATGTACGGAATCCGGTTACACCCCTGCCNAGGA 64

QY 600 CAGGGAAGCAGCAGAGATCAATCTGGCGCTTGCACACGCGGAATCTAGTGTTCAGG 659  
DB 65 CAGGGAAGCAGCAGAGATCAATCTGGCGCTTGCACACCTGGAATCTAGTGTTCAGG 124

QY 660 TTTCACCTAGATCAATGCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGGA 719  
DB 125 CTTTCAGGAAGATCAATGCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGCGGA 184

QY 720 GCGCTTTCTCATCAAGCTCCGGCCAGATGCCAATAGTGTGCTACCAAGATACCTTGGAAAT 779  
DB 185 GCGCTTTCTCATCAAGCTCAGACCTGACGTGAATAGCTCATACCAAGACACCTTGGAGT 244

QY 780 CTTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGGAATACTGTGTGAACATCATGC 839  
DB 245 CTTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGGAATACTGTGTGAACATCATGC 304

QY 840 CTTCTTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCCGCTCTTTAGCCGGG 899  
DB 305 CTTCTTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCCGCTCTTTAGCCGGG 364

QY 900 GTCTCATTTTCGGTTTCAGTGGTGGACTCAGAAAGCAGAGTTCTCGACTATGTTAAAGAAG 959  
DB 365 GTCTCTCTCCGGTTTGTAGTGGTGGACTCAGAAAGCAGAGTTCTCGACTATGTTAAAGAAG 424

QY 960 AGACATAAGAAGTGCATTTGAAAGGAAGCAGACAGATTCATTTATCCGGAGCT 1019  
DB 425 CGGCCCAAGAAGTGCATTTGAAAGGAAGCAGACAGATTCATTTATCCAGGAGCT 484

QY 1020 TGCTTCACAGCTTACAGAACTGAATTCGGAAGTGTGGAGCAGTCTCAGCAGACCAAG 1079  
DB 485 GTCTTCACAGCTTACAGAACTGAATTCGGAAGTGTGGAGCAGTCTCAGCAGACCAAG 544

QY 1080 CTTTACATTTGGAGAGTGCAGATTTCCAGGGGCCAGAGTCCGGCGGAGGAAGGA 1139  
DB 545 CTTTACATTTGGAGAGTGCAGATTTCCAGGGGCCAGAGTCCGGCGGAGGAAGGA 604

QY 1140 ACCGAAGTTTCGGCGGGAGCGGGTGCACCGAGCCCTGCGCCGAGGAGAGCC 1199  
DB 605 AACAAAAGCTGTACTTGGAGTTGGAGCTTCAACAAAGCCCGCCCTCCCTAAGAGCC 664

QY 1200 CGCGGTAAACAGCAGCGGAGCGGAGCCGCTCGCGGCCACGAGGA- - - - -GAGGA 1253  
DB 665 TCAGGAGCAAGCTGCAGATGGAAACCAACCGTGTGGCCCTGAGGAAGAAGAGGAGGA 724

QY 1254 GGAGTGTGTTAAGGATGAGCAGCAGAGTAAACCTCAGCCCGCCAGCCCAAGCAGG 1313  
DB 725 AGAGGTGCGCAAGGAGCGCATCCGCGCAAGTAACTCAGCCCGCCAGCCCAAGCAGG 784

QY 1314 CTCCCTGACTGGAGTCTCAGCTTTCCAGAGTGTCTGTGAATCTGCAGGGGGAGTGGC 1373  
DB 785 CTCCCTGACTGGTGTAGCTTCCAGCTTTCAGAGTGTCCATCAACTCAGAGGAGGAGCCG 844

QY 1374 CCTCGCAAGCTGACCTTCTCCCAACCTGAGCCCGCCAGCAGCAAGCAGGCTCTCCCTT 1433  
DB 845 CCAGCCCAAGTGTAGCTTCTCCCAACCTGAGTCTGTGACAAAGCAGGAGGCTCTCCCT 904

QY 1434 GATCAGCCCGCTGCTGAATGACAGCGCTTCCCGCCGAGCGAGTGAAGTGAAGGCGG 1493  
DB 905 GATCAGCCCGCTGCTGAAGCAGCAGCGTCCCAAGCAGCGAGGAGGAGGCGGCG 964

QY 1494 GAGGAAGATTTCCCAACTGATAAGCGTACTTATAGTGAAGAGTGTCTACCAAGGA 1553  
DB 965 GCGAAAGATTTCCCAACTGATAAGCGTACTTATAGTGAAGAGTGTCTACCAAGGA 1024

QY 1554 GCGAACATATCTGAAGGATCTCGAAGTTATCACTTCGTGTTTCAGAGCACAGTGAACAA 1613  
DB 1025 GCGCACGTACCTGAAGGAGCTTTGAGTCAATTCGTGTTTCAGAGCACAGTGAACAA 1084

QY 1614 AGAGGACGCCATGCGCGGAAGCAGCTGAAAAGTCTCATATATCCCGAAATTTGAACCTTTGCA 1673  
DB 1085 AGAGGACTCCATGCGCGGAAGCCTTGAAGAGTCTCATATTTCCCAATTTTGAACCTTTGCA 1144

QY 1674 CAAAATTCATATAATTTTCTCAAGGAAATGAGCAACAGCTTGCCTGTGGAGGCGG 1733  
DB 1145 CAAATTCATACCAATTTTCTGAAGGAAATGAGCAACAGCTTGCCTGTGGAGGCGG 1204

QY 1734 CTTCAATGCCCAAAATCA- - -GAGATTACCAAGAAATCGCGATGTCTGAAGAACAT 1790  
DB 1205 CTCAAATGCCCGAGCTCAGAGGAGATTACCAAGAAATCGCGGAGCTCATGCTGAAAACAT 1264

QY 1791 TCAGGCGATGAAGCAGCTGCGCGCTCACCTGTGGAAGCAGCAGCGAGCCCTTGAGAGCCCT 1850  
DB 1265 TCAGGCGATGAAGCAGCTGCGCGCCACCTGTGGAAGCAGCAGTGAAGCCCTTGAGAGCCCT 1324

QY 1851 GGAGAAATGAATCAAGAGCTCCCGGGGCTGAGAAATCTTCTGAGAGACTTTGAGCTGCA 1910  
DB 1325 GGAGAGCTCCATCAAGGCTCGCGCGCTGGAACATTTTCTGCCGAGACTTCGAGCTGCA 1384

QY 1911 GAAGTGTGTTTACCTACCGCTCAACACCTTCTCCTCGCGCCACTGCACCGCTCATGCA 1970  
DB 1385 GAAGTGTGTTTACCTGCGCTCAACACCTTCTCCTCGCGCCGCTGCACCGCTCATGCA 1444

QY 1971 CTACAGCAGGTCCTGGAGCGCTGTGCAAAACACACCGCGGAGCCGAGCTTCAG 2030  
DB 1445 CTATAGCATGCTCTGGAGAGCTGTGCAAGCAGCAGCCACCAACACCGAGCTTCAG 1504

QY 2031 GGAGTCCGAGCGCTTTGGCAGAGATCAGGAGATGTTGGCAGAGCTCCAGGTTACAT 2090  
DB 1505 GGAGTCCGAGCGCTGGCGGAGATCAGAGATGTTGGCAGAGCTGCATGTTACCAT 1564

QY 2091 GATCAAGATGGAAGATTTCCAGAACTGCACAACTCTCAAGAAAGATTTGATTTGGCATCA 2150  
DB 1565 GATCAAGATGGAAGATTTCCAGAACTGCAGAACTGCAGAAAGATCTCATCGGATCA 1624

QY 2151 CAATCTTGTGTTCCGGGAAGGAGTTTCATCGCTCTGGCAGCGCTCAGCAAGCTCTCGG 2210  
DB 1625 CAATCTTGTGATCCCGAGGAGGAGTTTCATCGCTCTGGCAGCGCTCAGCAAGCTCTCGG 1684

QY 2211 GAAGGGCTCAGCAGCGCATGTTCTTCTGTTCACAGAGTCTCTATACAGGCGG 2270  
DB 1685 GAAGGGCTCAGCAGCGCATGTTCTTCTGTTCACAGATGTTCTGCTATACAGCGG 1744

QY 2271 GGGGCTGAGCGCTCCAATCAGTTTAAAGTCCAGGCGAGCTCCCGCTCTATGGCATGAC 2330  
DB 1745 GGGAGTGAAGGATCTATATCAGTTTAAAGTCCAGGAGCTCCCACTCTATGGCATGAC 1804

QY 2331 GATTGAGGAGCGGAGACGA 2351  
DB 1805 GATCGAGGAGGTGAGGAGGA 1825

RESULT 3  
BC009153  
LOCUS BC009153 3719 bp mRNA linear ROD 07-AUG-2002  
DEFINITION Mus musculus, Similar to KIAA0793 gene product, clone MGC:6304  
IMAGE:2655209, mRNA, complete cds.  
ACCESSION BC009153  
VERSION BC009153.1 GI:14318718  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 3719)  
AUTHORS Strausberg,R.  
TITLE Direct Submission

JOURNAL Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunnarone, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK plate: 5 Row: m Column: 20.

FEATURES Location/Qualifiers

1..3719  
    /organism="Mus musculus"  
    /db\_xref="taxon:10090"  
    /map="FVB/N-3"  
    /clone="MGC:6304 IMAGE:2655209"  
    /tissue\_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."  
    /clone\_lib="NCI\_CGAP\_Mam2"  
    /lab\_host="DH10B"  
    /note="Vector: pCMV-SPORT6"  
    99..3296  
        /codon\_start=1  
        /product="similar to KIAA0793 gene product"  
        /protein\_id="AAH09153.1"  
        /db\_xref="gi:14318719"  
        /translation="MGTEGTVRALPTSGTRLGGTATGVSTLEPEQSLSPRMOEKHM RIRKVLDRKNAFLDIPEKQDQVLLTOVWHLNLECDYGLFKPKNVOSYWIWLEPM KPIIROVRKPNVLRVLAKEFPDPGLOEYTRYLEALQKRDLLERLITCNTANTA ALLSHLQSEIGDEYDLREHLKANLEYLPNOESLEKILDFHORHTGQTPAESDFQ KLEFLIKLHPVHPQIDQLEFLLSRDECKNFWKICVEYHTFRLSDQPKAKAVF FSRGSSFRYSGRTQKLVYVKGDMKRIPIYERRHSKRTSLHALTVLDPKOSVSFTD GLRTSASLSANLVFPYPPSSSLPPLNLDKSSSLVDPAVPKSTAERSGSPS SDDGPTOSALHGPVLPVLRPGFSDPSQPSLSKSHLSLCPQLQALSTAEOGAS PVLSPVLSGAGTARNDGEOEKHKHMPDEAFYFAKEILATERYLYKLEIVTYWFRS VLIKEEAMPALMALLFSNIDPVYEFHRGFLHEVEQRLALWEGPSSAHLKGDHQIRIGD ILLNMRLQKFTSYFRHDEVLTELEKATKCKLEAVYKEFELOKVYCLPNTFLFLL KPVORLHYRLLLSLRCAHYSPGHRDYADCHEALKAIPTVTELOQSLTRLENLQKLT ELORDLVGNELIAPGREFFIREGCLHLTKKGLQORMFLFSDMLLYTSKSVTGASHF RIRGLPLRLGVESSENEVSLHCFTIYAQKTIYVAASTRELEKEMQDLNRAIQ A KTIQDPSVLLGPPVYTRPRSSDESEEDGRNGRSGLENSQHRANTTHWCV YRNTSVSRADHASAVENSLGKYLKFKNSNGWKLWVYFNFLFFYKTHODDYPLA SLPLIGVSVPLEADSTHKDYVFKLPKSHVYFFRAESKYTFERWMDVIKRRSSPG RPPSPQCSHHSGLEAIREKEACPSPCLDKNL"

BASE COUNT 997 a 979 c 929 g 814 t  
ORIGIN

Query Match 31.0%; Score 728.8; DB 10; Length 3719;  
Best Local Similarity 58.5%; Pred. No. 4.8e-165;  
Matches 1374; Conservative 0; Mismatches 947; Indels 27; Gaps 5;

QY 10 AAACGTGCTCATCAAAATCCAGATGCTGGATGACACCCAGGACGATTTGAAGTTCA 69  
Db 222 AGCACATCGCATCAGATGAAGTCTGGACAGCACTGTAGAGTATTTGACATCGAG 281

QY 70 CAAAGAGCTCTGGAGGTGCTGCTGATGAGTTCAGTTCGAACCACTCAACCTCGTGGAA 129  
Db 282 CGAAGTGTGACGGCAGGTCTTACTGACTCAAGTGTGGAAGCATTGTAATCTCATTTGAA 341

QY 130 GGTGACTATTTTGGCTCGAGTTTCTCTGATCACAAAAAGATCACGGTGTGGTGGATCTC 189

Db 342 TGTGACTACTTTTGGTCTGGAGTTCAAGAATGTCCAGTCTCTACTGGATTGGTGTGAACCT 401  
QY 190 CTAATAACCCATTGTGAACAGATTAGAAGGCCAAAGCACGTTGTTGTTAAAGTTTGTGGTG 249  
Db 402 ATGAATAACCCATCATTTAGGCAAGTACGAAAGGCCAAATAATTCGCGTGTCTCGCTGGCAGTA 461  
QY 250 AAATTTCTTTCCGCTTGACACACACAACTCCAAAGAAAGAACTCACAAAGGTACCTGTTGCGG 309  
Db 462 AAATTTCTTTCCGCTTGACACACACAACTCCAAAGAAAGAACTCACAAAGGTACCTGTTGCGG 521  
QY 310 CTGCGAGTGAAGCAGGACTTGGCTCAAGCCAGTGTGACGTGTAATGACACACAGCGCAGCT 369  
Db 522 TTGCAACTCAAGAGAGACCTCTCTGGAGGAACGCTGTGACGCGCCACACACTGCGAGCC 581  
QY 370 CTTCTGATTTACACATTTGCAATCTGAGATTGGGAGTTTGGGATTTTGTGATGAAGCCTTGGACAGA 429  
Db 582 CTTCTCATATCCCACTTCTGCAGTCGGAATTCGGAGATTATGATGAACCTTGGATCGA 641  
QY 430 GAGCACTTAGCAAAAAATAATACATACCTCAGCAAGACGCACTAGAGGACAAAAATTCGTG 489  
Db 642 GAACACCTCAAGGCCAATGAATACTTGCCCAACCCAGGAGAAATCTCTAGAAAAAGATACTA 701  
QY 490 GAATTTCCACATAACACATTTGGACAAACACACAGCAGATCAGATTTCAGCTCTCTAGAG 549  
Db 702 GACTTCCATCAGAGGCACACGGGCCAGACTCTCTGCAGAGTCAAGATTTCAGAGGTGCTTTGAG 761  
QY 550 ATTGCCGCTGCGCTAGAGATGTATGGAATCCGTTGCAACCGGCCCAAGCACAGGGAAGCC 609  
Db 762 ATTGCAAGGAATCGAAATGTATGATGATCAGGTTTCACATGGCTTCTCAGCAGAGAAAGG 821  
QY 610 ACGAAGATCAATCTGGCCGCTTGCACACACGGAAATTTCTAGTGTTCAGGGTTTCTACTAAG 669  
Db 822 ACCAAGATTAACTAGCAGTTTCTCACATGGGTGCTCTGTTCCAGGTACCACCAA 881  
QY 670 ATCAATGCTTCAACTGGGCGCAAGGTGCGGAAGCTGTGCTTCAAGAGGAAGCGCTTTCTTC 729  
Db 882 ATCAATACCTTCAACTGGTCCAAAGTCCGTAAACTAAGCTTCAAGAGAAAAAGGTTCTT 941  
QY 730 ATCAAGCTCCGCGCAGATGCCAATAGTGTGCTACGAGGATACCTTGGAAATTCCTGATGGCC 789  
Db 942 ATCAAACTCCACCCCTGAGGTGCTACGGCCCTACCGAGACACGTTAGATTGTTGTTGGGT 1001  
QY 790 AGTCGGGATTTCTGCAAGTCCTTCTGGAAAACTCTGTGTGAACATCATGCTTCTTTTGA 849  
Db 1002 AGCAGAGATGAATGTAAGAATCTCTGGAAGATATGTGTGAGTACCATACCTTTTTTGA 1061  
QY 850 CTTTTTGAAGACCCCAACCAAGCCCAAGCCGCTCTTTTAGCGGGGGTCAATCATTT 909  
Db 1062 CTCTCTGACGAGCTTAAGCCAAAGGCGTGTCTTCTTTCAGCGGAGGCTCTCTCTTC 1121  
QY 910 CGGTTTCAGTGGTGGCAGTCAAGAGCAGGTTCTCGACTATGTTAAAGAGGAGGACATNAG 969  
Db 1122 AGATACATGGAAGAACTCAGAAACCAACTAGTAGATTATGTCAAGAGCGGTGAATGAAG 1181  
QY 970 AAGGTCGATTTGAAAGGAAGCACACAGCAAGATTC---ATTCTATCCGAGCCTTGCTTCA 1026  
Db 1182 AGAATTCATACGAAGACGCGCACAGTANAGCTCGGACATCTCTTCATGCTCTGACTGTA 1241  
QY 1027 CAGCCTTACAGAATGAATTTGGAAAGTGTGAGCAGTCTCTCAGCAGAGCACCAGCCTTACA 1086  
Db 1242 GATCTGCTTAAACAGAGCGCTCTCTTACCGGATGGCTTTCAGAACTTCTGCTCCCTGTCT 1301  
QY 1087 TTTGGAGAGGTGGCCAAATCTCCAGGGGCCAGAGCTGCGCGGAGGAGAAAGAACCGAAG 1146  
Db 1302 TCAGCAAAATGTCTCTTTTATCC-----ACCCCTAGTCTTCTCTTTGTCCCTCC 1351  
QY 1147 GTTTCCGCGGGGCGCGGGTCTCGACCCGAGCCTGCGCCGAGGAGAAAGCCCGCGGT 1206  
Db 1352 TGGTCTGCCCAATTTGAAGGACAGCAGCAGCTCCCTTGTGGATCCCGAGGCTCCCGTCA 1411  
QY 1207 ACAAAGCAGCGGCGGAGCGCCCTCGGCGCCCAACCGAGGAAGAGGAGGTGCTTTAAG 1266

Db 1412 CAAGACACAGCAGCAGAGAGGAGGAGCGGAGCCATCATC--ATCTGAGCGGCCCCAGCAC 1469

QY 1267 GATAGACCCAGCAGATAACCTCAGCCCCGCGAGCCAAAGCACAGGCTCCCTGACTGGC 1326

Db 1470 CAGTCGGCCCATCTCCCTGGACCCCGGTGCTCCCGGCTGGTTCCTATGAT 1529

QY 1327 AGTCCTCAGCTTTCCGAGCTGTCTGTAACTCGCAGGGGGAGTGGCCCTGCCAACGTG 1386

Db 1530 AGTCCTCAGCTTTCCCTCCAGCTCGAGCCACCTGAGCTGTGCCCCTGAGCTTCAG 1589

QY 1387 ACCTTGCTCCCAACTGAGCCCCGACACCAAGAGGCGCTCTCCCTTGATCAGCCCGCTG 1446

Db 1590 GCCGG-----CTGAGTACAGCTGAGCAGGGTGTGATCCCTGTGCTCAGCCCTGTG 1640

QY 1447 CTGAATGACCGGCTGCCCGCGAGGACGATGAGGATGAGGCGCGGAGGAGATTC 1506

Db 1641 CTCAGTGGCGTGTACAGCCAGGATGACACCAAGAAAGACAGAAACAGCACATG 1700

QY 1507 CCAACTGATAAAGCGTACTTCATAGCTAAGGAAGTGTCTACCACCGAGCGACATATCTG 1566

Db 1701 CCAGAAGATGAGCCCTATTTCATAGCCAAAGGAGATTCTCGCTACAGAACGACCTATCTG 1760

QY 1567 AAGGATCTCGAAGTATACATCTCGTGGTTTCAGACACAGTGTGAGAAAGAGACGCCATG 1626

Db 1761 AAGGATTTAGAAGTTATCACTGTGTGGTTCGCGAGCGTGTGATCAAGGAGGAGGCCATG 1820

QY 1627 CCGGAGCAGCTGAAAGTCTCATATTCCGGAATTTGAACTTTGCACAAATTTTCATCT 1686

Db 1821 CCTGACGCGCTGATGGCCCTGCTTTTCTCAACATTTGATCCAGTCTACGAGTTCACACAGA 1880

QY 1687 AATTTTCTCAAGAAATGAGCAAGCACTTGCCCTGTGGAGGCGGCTCAAAATGCCCA- 1745

Db 1881 GCGTTCCTCACGAGTGGAAACAGAGGCTGGCACTCTGGGAAGGCGCTCCAGTCCCCAC 1940

QY 1746 --AATCAGAGATTACAAAGAAATCGCGATGTCTAGTGTGAAGAAATTCAGGGCATGAAG 1803

Db 1941 TTAAGAGGTGATCACCAGCAGTCCGGGACATCTCTCCACGAAACATGCGTCAGTTAAAG 2000

QY 1804 CACCTGGCGCTCACCTGTGGAAGCACAGCGAGGCGCTTGGAGGCGCTGGAGAAATGGAATC 1863

Db 2001 GAATTTACTAGTACTTCCAAAGACACAGATGAGGTCTTAACAGAACTGGAAAGGCCACA 2060

QY 1864 AAGAGCTCCCGCGCTGGAGAACTTTCGACAGACTTTGAGCTGCAGAAAGTGTGTAC 1923

Db 2061 AAACACTGTAAAGAGCTGGAGCAGCTTACAAAGAGTTTGAAGTCCAAAGAGTCTGCTAC 2120

QY 1924 CTACCGCTCAACACCTTCTCTCTGGGCGCAGTGCACGCGCTCATGCATCAAGCAGGTC 1983

Db 2121 CTGCGCTCTCAACACATCTCTGCTGAAGCGCGTCCAGAGGCTAGTCCACTACCGTCTGCTG 2180

QY 1984 CTGGAGCGGCTGTGCAAAACACCACCGCGGAGCCAGCGGCTTACAGGAGTCCCGGAGCC 2043

Db 2181 CTGAGCGGCTGTGTCTACTACTCTCTCTGGGACCGGAGTATGCCGACTGCCCATGAG 2240

QY 2044 GCTTTGGCAGAGATCACGGAGATGTTGGCCACAGCTCCACGCTACGATGATCAAGATGGAG 2103

Db 2241 GCACCTGAAGGCCATCACAGAAGTGAACCCAGCTCCAGCAAGGCTTACCCGGCTGGA 2300

QY 2104 AATTTCCAGAAGCTGCAGAACTCAAGAAAGATTGATTGGCATTTGACAACTTTGTGGTT 2163

Db 2301 AACCTACAGAAATTCACGAGAGCTACAGGAGACCTTGGTGTGTAAGAAACCTCATTTCT 2360

QY 2164 CCGGAAGGAGTTCATCTGCTGGGACCGCTCACAGCTCTCGGGAAGGGGCTCCAG 2223

Db 2361 CTTGGAGGAGATTATCCGTGAGGCTGCTCTGCAACAGCTCACCAGAAAGGGCGCTGGA 2420

QY 2224 CAGCGCATGTTCTCTCTTCAACAGCTCTCTGCTATACACGAGCGGGGCTGACGGCC 2283

Db 2421 CAGAGATGTTTTTCTGTCTCAGATATGTTTACTGTATACAAAGCAAAAGTGTACACAGA 2480

QY 2284 TCCATCATGTTTAAAGTCCACGGGAGCTCCCGCTCTATGCCATCAGCATTCAGGAGAGC 2343

Db 2481 GCCAGTATTTCCGGATCCGTGCTTCTTCCACTCCGTGGCATGCTGTTAGAAAGAAAGT 2540

QY 2344 GAAGACGA 2351  
|| ||  
Db 2541 GAGATGA 2548

RESULT 4  
AB018336  
LOCUS AB018336 3997 bp mRNA linear PRI 16-JUN-1999  
DEFINITION Homo sapiens mRNA for KIAA0793 protein, complete cds.  
ACCESSION AB018336  
VERSION AB018336.1 GI:3882306  
KEYWORDS  
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone\_lib:pBluescriptII  
SK plus clone:hk05692.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (sites)  
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,  
Tanaka,A., Kotani,H., Nomura,N. and Ohara.O.  
Prediction of the coding sequences of unidentified human genes. XI.  
The complete sequences of 100 new cDNA clones from brain which code  
for large proteins in vitro  
DNA Res. 5 (5), 277-286 (1998)  
99087487

REFERENCE  
2 (bases 1 to 3997)  
Ohara.O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.  
Direct Submission  
Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute,  
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba  
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913,  
Fax: +81-438-52-3914)

FEATURES  
Location/Qualifiers  
1..3997  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hk05692"  
/sex="male"  
/tissue\_type="brain"  
/clone\_lib="pBluescriptII SK plus"  
/dev\_stage="adult"  
1..3997  
/gene="KIAA0793"  
118..3282  
/gene="KIAA0793"  
/codon\_start=1  
/product="KIAA0793 protein"  
/protein\_id="BAA34513.1"  
/db\_xref="GI:3882307"  
/translation="MGIEGT RVLTAGMRLGAQTPGVSTLEPGOTLLPRMQEKLH  
HLRVKLLDNTMEIFDIEPKCDGQVLLTQVWKRNLNVECDYFGMEFQNTQSWINLEPM  
KPIILRIIRRNPNVRLAVKFFPPDQQLQEEYTRYLFAQLKRDLLERLTCADTTA  
ALSTSHLQSEIGDYDETLDREHLKVNXYLPQQQCLEKLEFHOKHVGQTPAESDFQ  
VLEATKLEMYGIRFHMASDREGTKIQLAVSHMGVLVFGQTKINTFNWSKVRKLSFK  
RKRELIKLHPVHGYPQDTLEFLGSDCKENWKICVEYHTFFRLDPOPKAKAVF  
FSGSSFRYSCRTQKLVDFKDSMKRIPYERRSHKTHTSVRALTADLPKQTSFPE  
GLRTPPSSANAEYSLSPSTLVPSGLPEFKDSSSLTDPOQVYVKSPAAERRSGAVA  
GGDTPSAQPLGFPALQPGELSTSPQSPSSRSKPLSPAFQVPLGPAEQSSPL  
LSPVSDAGAGGMDCEPRHKRVPADEAYFIVKEILATERYTKDLEVITVWFRSAV  
KDAMPATLMTLLFSDNIDPIYEFHGFLEVEQRLALWEGSKAHTKGSORIQDILL  
RNMQLKEFTSYFORHDEVTLEKATCKCKLEAVYKEFELQKVCYLPNLTLLKPI  
ORLHYRLLLRLLCGHYSPGHHDYADCHDKAITEVTTTLQHLIRLENLQKTELO  
RDLVGIENLTAPGREFIRESCLLHLTKKGLQORMFFLESDMLLTKSGVAGTSFPIR  
GLPLQLMVEESDNESVPHCFITYAAQKTIIVAASTRLEKEKMKDLDLNSAIQAAS  
GGDTPALPGRTVCTRPSPNEVLSQESDEDDARGVRSLEGGHQRANTMTVCWY  
RNTSVSRADHSAAVENOLSGVLLRKFKNHGKQLWVFTNFCFLFFKTHODDPLAS  
LPLLGYSVIPREADGIHKDYVFKLQFKSHVYFRAESKVTFRWMEVIQGAASSAGR  
APSTVQDGPQPPSSGLEGMGRKEP"

gene  
CDS  
1..3997  
/gene="KIAA0793"  
118..3282  
/gene="KIAA0793"  
/codon\_start=1  
/product="KIAA0793 protein"  
/protein\_id="BAA34513.1"  
/db\_xref="GI:3882307"

BASE COUNT 998 a 1132 c 1068 g 799 t  
ORIGIN  
Query Match 30.6%; Score 719.6; DB 9; Length 3997;

Best Local Similarity 58.3%; Pred. No. 8.1e-163; Matches 1372; Conservative 0; Mismatches 944; Indels 36; Gaps 5;			
QY	6	AGGAAATCGTCTCCATCAAAATCAGATCTGGATGATCAGCCAGGCGCATTTTCAAGT	65
Db	237	AGAGAGCAGCTGACCTCAGAGTAAGCTGTGGACACACCATGGAATATTGACAT	296
QY	66	TCACAAAAGAGCTCTGGGAAGGTGCTGGTGGATGAGTGTGCAACCACTCAACCTCGT	125
Db	297	TGAGCCTAAATGCGATGGCCAGGTATTACTGACACAAGTGTGGAACGCTTTAAACCTGGT	356
QY	126	GGAGGTGACTATTTTGGCTCGAGTTTCTGTATCAAAAAGATCACGGTGTGGCTGGA	185
Db	357	AGAATGTGACTACTTCCGGATGGAGTTTCAAAATACTCACTGCTACTGGATTGGCTTGA	416
QY	186	TCCTCTAAATCCCATTTGTGAACAGATTAGAAAGGCCAAAGCGTTTGTAAAGTTTGT	245
Db	417	ACCTATGAATCCCATCATTTAGGCAATATGAGAGGCCAAAGATGTGGCTTCGCCCTAGC	476
QY	246	GGTGAATTTCTTCCGCTGACACACACAACTCCAAGAGAACTCACAAGTACCTGTT	305
Db	477	TGTAAAAATTTTCCACCTGATCTCGTCACTTACAAAGAAATATACAAGATACCTGTT	536
QY	306	CGCGCTGCAGGTGAACAGAGCTTGCTCAAGCAGGTTGACGTGTAAATGACACACAGCC	365
Db	537	TGCTTTGCAACTTAAGAGAGACCTGCTGGAAGAGCGTTTGACGTGTGCTGACACACAGC	596
QY	366	AGCTCTTTGATTTCACATTTGCAATCTGAGATTTGGGATTTTGTGAAGCCTTGA	425
Db	597	GGCCTTCTCAGTCCCATCTCTCGAGTCGGAAATAGGAGATTAGTAACAGCTGGA	556
QY	426	CAGAGAGCACTTAGCAAAAATAATATACATACCTCAGCAAGCGCACTAGAGGACAAAAT	485
Db	657	CCGAGAGCACTCAAAGTGAACAGATTTTGGCTGSCCAGCAGCACTGCTTGAGAAT	716
QY	486	CGTGGAAATTCACCATTAACCATTTGGACAACACACAGCAGAAATCAGATTTCCAGCTCT	545
Db	717	ACTAGAATTTCCATCAGAACACGTGGGCCAGACACCTGCTGATCGGATTTCCAGGTGCT	776
QY	546	AGAGATTGCGCTCGGCTAGAGATGTATGAATCGGTTGACCGGCCCAAGSACAGGCA	605
Db	777	CGAATTTGCTCAAGTTGGAATGTACGGCATCAGATTTCAATGGCTTCGACAGGGA	836
QY	606	AGGCAGAGATCAATCTCGCGCTTGCCCAACACGCGGAATTTAGTGTTCAGGGTTTCAC	665
Db	837	AGGAACCAAGATTTCAACTGGCAGTTTCCACATGGGTGTACTCGTGTTCAGGGCCACAC	896
QY	666	TAGATCAATGCTTCAACTGGCCCAAGGTGCGGAAGCTGAGCTTCAAGAGAGCGCTT	725
Db	897	CAAAATCAACACTTTCACTGGTCCAAAGTCCGTAAACTAAAGCTTCAAGAGAAAAGATT	956
QY	726	TCTCATCAAGCTCCGGCCAGATGCCAATAGTGGTACCAGGATACCTTGGAAATTCCTGAT	785
Db	957	TCATTCAAACTTCATCCAGAGGTTTACAGGCTTACAGGACACATTAAGATTTTGT	1016
QY	786	GGCAGTCGGGATTTCTGCAATCTCTTGCAACACGCGGAATTTAGTGTGGAACATCATGCTTCT	845
Db	1017	GGTAGTAGAGATGAATGTAAAGATTTCTGGAAGATTTGTGGAGTATCACACCTTTT	1076
QY	846	TAGACTTTTGAAGAGCCCAACCAAGCCCAAGCCCGTCTCTTTAGCCGGGGTCAATC	905
Db	1077	TGACTTTTGGACCACTTAAGCCAAAGCAAAAGCGGCTTCTTTCAGCCGGGGTCCCTC	1136
QY	906	ATTTCGTTTCACTGGTGGCAGTTCAGAACAGGTTCCTGCACTATGTTTAAAGAGGAGACA	965
Db	1137	CTTCAGATACAGTGGAGACACTCAGAAACACTAGTAGATTTATTTCAAGACAGTGGAAAT	1196
QY	966	TAAAGAGTGCAGTTTGAAGGAAGCAGCAGCAAGATTTCAATTCATCCGAGCGCTTGTTC	1025
Db	1197	GAAGAGAATTTCCATATGAAGAAGGCACAGCAAGACCCACACGTCC-----	1242
QY	1026	ACAGCCTCAGAACTGAATTCGGAAGTGTGGAGAGTGTGAGAGTGTGAGAGACACAGCTTAC	1085



QY	786	GGCCAGTCGGGATTTCTGGAAGTCTTCTTGAAATCTGTGTGAACATCATGCTCTTT	845
Db	1025	GGGTAGTAGAGATGAATAGTAAGAACTTCTGGAAGTGTGTGAGTATCACACCTTTT	1084
QY	846	TAGACTTTTGAAGAGCCCAAGCCCAAGCCGCTCTTTAGCCGGGGTCATC	905
Db	1085	TAGACTTTTGGACCACTAAGCAAAAGCAAAAGCGTCTTCTAGCCGGGGCTCTC	1144
QY	906	ATTTCGGTTCAGGTGCGGACTCAGAGCAGGTTCTCGACTATGTTAAAGAGGAGACA	965
Db	1145	CTTCAGATACAGTGGAGAACTCAGAACTAGTAGATTAATTCAGACAGTGAAT	1204
QY	966	TAGAAGTGCAGTTGAAAGGAGCAGCAGAGTATTCATTCATCGGAGCCCTTGCTTC	1025
Db	1205	GAAGAGAAATTCATATGAAAGAGGACAGCAGAACCCACACGCTCGCTTG-----	1255
QY	1026	ACAGCCTACAGAACTGAATTCGGAAGTCTGGAGCAGCTCTCAGCAGACGACCTTAC	1085
Db	1256	-----AGCTCTGACTCGACCTACCAAAACAGAGCATCTCATTTCCCGAGGATT	1306
QY	1086	ATTTCGGAAGGTGCGGAATCTCCAGGGGCCAGAGCTGCCGCGAGGAAAGAACCGAA	1145
Db	1307	GAGGACTCTCTCCCTCCCTCTCAGCGAATGCTTTTACTCGCTCTCTCTCCCTCACTCT	1366
QY	1146	GGTTTCGCGCGGAGCGCGGGTTCGACCCGAGCCCTGCGCCGAGGAGAACCCCGGGG	1205
Db	1367	GGTCCCTCTGGCTGCGCAGATTTAAGGACAGCAGCAGCTCCCTCACAGATCCCGAGT	1426
QY	1206	TAAACAGCAGCGGAGCGGCGCTCGGCGCCACGAGGAGAGGAGGAGG-----TCGT	1262
Db	1427	TTCTAGCTCAAGTCCAGCTCGACAGGCGCAGTGGAGCAGTGGCTGGAGGCCCGGA	1486
QY	1263	TAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCGCGCAGCCAGCAGCAGCTCCCTGC	1322
Db	1487	CACACCATCGGCCAGCCCTCGGCGCCCGCAGCTCCAGCCTGCTCCAGGCTTTCCAC	1546
QY	1323	TGGCAGTCTCTACCTTTCCAGAGTGTGTGTAAGTCTGCGAGGGGGAGTGGCCCTGCA	1382
Db	1547	GAAGAGTCTCTCAGCTTTCTCCCTCCAGCGGAGAGCCCTGAGTCTGAGCCCTGCATT	1606
QY	1383	CGTGACCTTCTCCCAACTGAGCCCGACACCAAGCAGGCTCTCCCTTGATCAGCCC	1442
Db	1607	TCAGTGTCCCTT-----TGGGCCCCAGCTGAACAGGGCTCATCCCCACTCTCTGAGCCC	1657
QY	1443	GCTGCTGAATGACAGGCTTGCCTCGGAGCGAGCAGTATGAGTATGAGCGCGGAGGAG	1502
Db	1658	TGTCTCAGTGATGCTGGCGGAGCGGGATGGACTCGAG-----GAGCCACAGACAGCG	1714
QY	1503	ATTCCCACTGATAAGCGTACTTCTATAGTAAGGAGTGTCTACACCGAGGAGACATA	1562
Db	1715	CGTGCCTGCAGAGGAGGCTTCTATAGTCAAAGAGATTCTCGCTACAGAGAACATA	1774
QY	1563	TCGTGAAGGATCTGAAATATCACTTCGTGTTTTCAGGCACAGTGAAGAGAGCGC	1622
Db	1775	CTCAAGGATTTAAGAGTATTACCGTGTGTTCCGAGCGCAGTGGTGAAGGAGGCGC	1834
QY	1623	CATCGCGGAGACCTGAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTTCA	1682
Db	1835	CATGCTCGGACTCTGATGAGCTGTCTTCTCAACATCGATCCCATCTATGAGTTCCA	1894
QY	1683	TACTAATTTCTCAGGAATTTGAGCAAGCATTGCGCTGTGGAGCGGCTCAATGC	1742
Db	1895	CAGAGGCTTCTCGGAGGTTGAGCAGAGGCTGGCACTCTGGGAAGGGCCCTCCAAAGC	1954
QY	1743	CCAAATCAGAG---ATTACCAAGATTCGCGATGTCATGCTCAAGAACATTC	1792
Db	1955	CCACAAAAGGAGCTCATCAAGAAATCGGGGACATCTCTGCTCAGGAACATGC	2007

RESULT 6  
BC004009 2632 bp mRNA linear ROD 07-AUG-2002  
LOCUS Mus musculus, clone IMAGE:3493093, partial cds.  
DEFINITION

ACCESSION	BC004009
VERSION	BC004009.1
KEYWORDS	GI:13278387
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 2632)
TITLE	Strausberg, R.
JOURNAL	Direct Submission
REMARK	Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabbs-r@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
FEATURES	Location/Qualifiers 1..2632 /organism="Mus musculus" /db_xref="taxon:10090" /map="CZCH II" /clone="IMAGE:3493093" /tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV." /clone_lib="NCI_CGAP_Lu29" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" <1..1263 /codon_start=1 /product="Unknown (protein for IMAGE:3493093)" /protein_id="AAH04009.1" /db_xref="GI:13278387" /translation="VLMKNIOGMKHLAAHLWKHSEALEALETSSIKGRRLLEHFCRDFE LOKCYLPLNTFLRLPLHRLMHYKHVLEBLCCKHHPNHDRCRAALAEITFEMVAQL HGTMLKMFQKLEKHLKDLIDGLNLIIPGREFIRGLSKLSKGLQRMFLFNDR LLYTSRGLTASNOFKVHGLPLYGMTIEESEEWGVPHCLTIRGQRSTIIVASRSR MEKMDIQMADLAEKSNPTPELLASPPDNKSPDEATADQESDLSASRTSLE RQAPRGNTMHWCHRSSTVSVDFTSAVENQLSNLIRKFNKSNWGLKLVVFTNF CLFFYKHQDSHPLASLPLGLSLTIPSENIHKDYVFKLHKSHVYVYFRAESEVTF ERWMEVIRSATSSASRAHLHSHKSHLY"
BASE COUNT	632 a 720 c 671 g 609 t
ORIGIN	Query Match 20.4%; Score 478.8; DB 10; Length 2632; Best Local Similarity 89.3%; Pred. No. 1.2e-104; Matches 516; Conservative 0; Mismatches 62; Indels 0; Gaps 0; QY 1774 GTCATGCTGAAGAACATTCAGGGCATGAAGCACCTCGGCGCTCACCTGTGGAAGCAGAC 1833 Db 1 GTCATGCTGAAGAACATTCAGGGCATGAAGCACCTCGGCGCTCACCTGTGGAAGCAGAC 60 QY 1834 GAGGCTTGGAGCCCTGGAGATGAATGAACAGAGCTCCCGGGCTGGAGACTTCTGC 1893 Db 61 GAGGCTTGGAGCCCTGGAGACTTCCATCAGGGCTCGCGGCTGGAGAACATTTCTGC 120



```
QY 1894 AGAGACTTTGAGCTGCAGAGAGTGTGTTACCTACCGCTCAACACCTTCCTCTCGGCGCA 1953
Db 121 CGAGACTTCGAGCTGCAGAGAGTGTGTTACCTACCGCTCAACACCTTCCTCTCGGCGCG 180
QY 1954 CTGACCGGCTCATGACACTACAGCAGAGTCCCTGGAGCGCTGTGCAAAACACACCGCGC 2013
Db 181 CTGACCGGCTCATGACACTATAAGCATGTCTTGAGAGGCTGTGCAAGCACCACCCACCA 240
QY 2014 AGCCACCGGCTTCAGGAGCTCCGAGCGCTTTGGCAGAGATACGAGAGTGTGGCA 2073
Db 241 AACACCGGCTTCAGGAGCTGCAGAGCTCGCTCGCGGAGATACAGAGATGTGGCC 300
QY 2074 CAGCTCCACGGTACGATGATCAAGATGAGAAATTTCCAGAAGCTGCACGAACCTCAAGAA 2133
Db 301 CAGCTGCACGGGACCATGATCAAGATGAGAACTTCCAGAAGCTGCATGAGCTCAAGAAA 360
QY 2134 GATTTGATGGCATTCACAACTCTTGTTGCTCGGGAAGGAGTTTCATCGCTGGGCAGC 2193
Db 361 GATCTGATCGGCATTCACAACTCTTGATGCCAGGAAGGAGTTTCATCGCTGGGCAGC 420
QY 2194 CTGAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGAGCTC 2253
Db 421 CTGAGCAAGCTCTCGGGAAGGGCTTCAGCAGCGCATGTTCTTCTGTTCAACGATGTC 480
QY 2254 CTGCTATACAGCGCGGGGCTGACGGCTCCAAATCAGTCTTAAAGTCCACGGGCGAGCTC 2313
Db 481 TTGCTGTATACCGCGGGGCTGACAGCATCTAATCAGTCTTAAAGTCCACGAGCGCTC 540
QY 2314 CCCTCTATGGCATGACGATTGAGGAGAGCGGAGAGCA 2351
Db 541 CCATCTATGGCATGACGATCGAGGAGAGTGAGGAGGA 578

RESULT 7
HSM802608 2888 bp mRNA linear PRI 23-MAR-2000
LOCUS Homo sapiens mRNA: cDNA DKFp762P046 (from clone DKFp762P046).
DEFINITION AL161984
ACCESSION AL161984
VERSION AL161984.1 GI:7328120
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Bloeker,H., Boeher,M., Brandt,P., Mewes,H.W., Weil,B. and
Wiemann,S.
JOURNAL Direct Submission
COMMENT Submitted (15-MAR-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFp762P046) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES
Location/Qualifiers
source
1..2888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="DKFp762P046"
/tissue_type="melanoma (MeWo cell line)"
/clone_lib="762 (synonym: hmel2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal 2852..2857
polyA_site 2859
BASE COUNT 879 a 592 c 620 g 797 t
```

```
ORIGIN
Query Match 14.6%; Score 342.8; DB 9; Length 2888;
Best Local Similarity 68.2%; Pred. No. 8,9e-72;
Matches 476; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 295 AGGTACCTGTTCCGGCTGCAGTGAAGCAGAGGACTTGGCTCAAGGCGAGTTGAGCTGTAAAT 354
Db 123 AGGTATCTTTTACTCTTCAAAATAAAGAGGATTTGGCTCTAGGAAGGCTTCCATGCACAT 182
QY 355 GACACAGCGCAGCTCTCTTTGATTTACACATATGTCGAATCTGAGATTTGGGATTTGAT 414
Db 183 GAACAATGTGACAGCGTTGATGTATCTACATCTTTACAAATCAGAACTTGAGACTTTTCA 242
QY 415 GAAGCTTGGCAGAGAGCAGCTTAGCAAAAATAATACATACCTCAGCAAGCAGCACTA 474
Db 243 GAAGAAACAGATAGAGAGCATCTGGCAAACTCGGTACTTTACCAACCCAGACTGTTTA 302
QY 475 GAGGACAAATCGTGGAAATTTCCACATACACCATTTGGACAAACACACAGCAATCAGAT 534
Db 303 GAGGCAAGATCATGCACCTTTCATCAGAGCAGATTTGGCAGGAGCCAGCTGAATCTGAC 362
QY 535 TTCCAGCTCCTAGAGATTTGCCGCTGGCTAGAGATGTATGGAATCGGTTGACCGCGCC 594
Db 363 ATTCTGCTACTGGACATAGCAAGAGCTGGATATGTATGGCATCAGGCTCACCCGCC 422
QY 595 AAGGACAGGGAAGGACGAGATCAATCTGGCCGTTGCCAACACGCGGAATCTAGTGT 654
Db 423 AGTGATGTTGAAGGATGCGAGATTCACCTGGCTGTTCTGCATCGGAGTACTGGTGT 482
QY 655 CAGGGTTTCACTAAGATCAATGCTTCACTGGGCAAGTTCGGAAGTCTGCTGAAATCTGTG 714
Db 483 CGGGAAATACAAAGATCAATCTTTAACTGGCTAAATCCGCAAGTTGAGTTTAA 542
QY 715 AGAAGCGCTTTCATCAAGCTCCGCGCAGATGCCAATAGTCGTACAGGATACCTTG 774
Db 543 AGAAAGCATTTTCTCATCAAACTTCATGCAATATCTTGGTGTGTGCAAGGATACCTTG 602
QY 775 GAATTCCTCATGCGCAGTCGGGATTTCTCAAGTCTCTCTGAAATCTGTCTGAAACAT 834
Db 603 GAGTTCACCATGCCAGCGCAGATGCTGCAAGGCTTCTGGAAGACTTGTGTGGAATAC 662
QY 835 CATGCTCTTTAGACTTTTGAAGAGCCCAACCAAGCCCAAGCCCTCTCTTCTTAC 894
Db 663 CATGCTCTTTCAGGCTTTTCGGAAGAGCCCAAAATCAAAAGCCCAAGCCCTCTCTG 722
QY 895 CGGGGTCATCATTTTCGGTTCAGTGGTGGACTCAGAACGAGGTTCTCGACTATGT 954
Db 723 AAGGTTCCAGTTTCCGCTATAGTGGACGACCCCAAGGCAACTTTTGAATATGGGAGA 782
QY 955 GAAGGAGGACATAAGAGGTGCAGTTTGAAGGAAGCA 992
Db 783 AAGGAGGCTCAAGAGCTTGCCATTTGAAGGAAGCA 820

RESULT 8
BC027077 2431 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, clone IMAGE:5010682, mRNA, partial cds.
DEFINITION BC027077
ACCESSION BC027077
VERSION BC027077.1 GI:20071584
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 2431)
JOURNAL Direct Submission
Submitted (04-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
```

REMARK COMMENT	<p>NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>  Contact: MGC help desk  Email: <a href="mailto:cgaps-remail.nih.gov">cgaps-remail.nih.gov</a>  Tissue Procurement: Gilbert Smith, Ph.D.  cDNA Library Preparation: Life Technologies, Inc.  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  Center code: BCM-HGSC  Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a>  Contact: <a href="mailto:amg@bcm.tmc.edu">amg@bcm.tmc.edu</a>  Guntaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kovis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.</p>									
FEATURES	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>  Series: IRAK Plate: 45 Row: C Column: 12  This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.</p>									
source	<p>Location/Qualifiers  1. .2431  /organism="Mus musculus"  /db_xref="taxon:10090"  /map="FVB/N-3"  /clone="IMAGE:5010682"  /tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."  /clone_lib="NCI_CGAP_Mam2"  /lab_host="DH10B"  /notes="Vector: pCMV-SPORT6"  &lt;1. .1065  /codon_start=1  /product="Unknown (protein for IMAGE:5010682)"  /protein_id="AAH27077.1"  /db_xref="GI:20071585"  /translation="YKHVLERLKHPHPNHADFRDCRAALAEITEMVAQLHGTMIKME  NFQKHELKDLIGIDNLVTPREFIRLAGLSKLSGKGLQQRFFFLNDVLLVTSRGL  TASQPKVHGQLPYGMTIESEEWGVPHCLTLRGORSIIVAASRSSEMEKMWGDI  OMAIDLAERKSGPTPELLASSPPDKSPDEATAADOESDLSASRTSLERQAPHRGI  TMVYQWHRSTSVSMVDFSAVENQLSNLLRKFKNNGWQKLWVTFNFCFLFFYKSH  QDSHPASLPLLSHTTPTSESNINHKDIVFKLHKSHVYFRAESEITFERWMEVIR  SATSSAAHILSHKRESHLY"</p>									
CDS	<p>581 a 659 c 614 g 577 t</p>									
BASE COUNT	<p>581 a 659 c 614 g 577 t</p>									
ORIGIN	<p>Query Match 13.3%; Score 312.8; DB 10; Length 2431;  Best Local Similarity 88.9%; Pred. No. 1.6e-64;  Matches 338; Conservative 0; Mismatches 42; Indels 0; Gaps 0;</p>									
QY	1972	TACAAGCAGGTCTCTGGAGCGGCTGTGCAACACACACCGCGGAGCCACGCGACTTCAGG	2031							
Db	1	TATAAAGCATGCTCTGGAGAGGCTGTGCAAGCACACCCACCAACACGCGGACTTCAGG	60							
QY	2032	GACTGCCGAGCCGCTTTGGCAGAGATCACGAGATGGTGGCACAGCTCCACGGTACGATG	2091							
Db	61	GACTGCAGAGTGGCTGGCGGAGATCACAGAGATGTTGGCACAGCTGCATGGTACCATG	120							
QY	2092	ATCAAGATGAGATTTCCAGAACTCCAGAACTCAAGAAAGATTTGATTGGCAATTGAC	2151							
Db	121	ATCAAGATGGAGAACTTCCAGAACTCATGAGCTCAAGAAAGATCTGATCGGCATTGAC	180							
QY	2152	AATCTCTGTGTCGGGAAGGAGTTTCATCCGTCGTGGCAGCCTCAGCAAGCTCTCGGG	2211							
Db	181	AATCTGTGACCCGAAGAGGAGTTTCATCCGCTGGCAGCCTCAGCAAGCTCTCGGG	240							
QY	2212	AAGGGGCTCCAGACGCGATGTTCTTCTGTTCAACGACGTCCTGCTATACAGAGCGG	2271							
Db	241	AAGGGGCTTCAGACGCGATGTTCTTCTGTTCAACGATGTTCTGCTATACAGAGCGG	300							
QY	2272	GGGTGACGGGCTCCAATAGTTTAAAGTCCAGGGCAGCTCCCGCTCTATGGCATGAC	2331							

Db	301	GGACTGAGCGCATCTAATCAGTTTAAAGTCCAGGACAGCTCCCACTCTATGGCATGACG	360
Qy	2332	ATTGAGGAGCGGAGGACGA	2351
Db	361	ATCGAGGAGAGTGAGGAGGA	380
RESULT 9			
LOCUS	AL161896		
DEFINITION	Human DNA sequence from clone RP11-261P24 on chromosome 13,		
ACCESSION	AL161896		
VERSION	ALI61896.16	GI:12330752	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 96183)		
JOURNAL	Phillimore,B.		
COMMENT	Direct Submission		
	Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,		
	CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk		
	Requests: clonerequest@sanger.ac.uk		
	On Jan 22, 2001 this sequence version replaced gi:12329254.		
	During sequence assembly data is compared from overlapping clones.		
	Where differences are found these are annotated as variations		
	together with a note of the overlapping clone name. Note that the		
	variation annotation may not be found in the sequence submission		
	corresponding to the overlapping clone, as we submit sequences with		
	only a small overlap as described above.		
	This sequence has been finished according to sequence map criteria		
	as follows. An attempt is made to resolve all sequencing problems,		
	such as compressions and repeats, but not necessarily within known		
	annotated repeat sequence elements. Where the sequence is		
	ambiguous, there is an annotation using the 'unsure' feature key.		
	The following abbreviations are used to associate primary accession		
	numbers given in the feature table with their source databases:		
	Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information		
	on the WORMPEP database can be found at		
	http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence		
	was generated from part of bacterial clone contigs of human		
	chromosome 13, constructed by the Sanger Centre Chromosome 13		
	Mapping Group. Further information can be found at		
	http://www.sanger.ac.uk/HGP/Chr13		
	RP11-261P24 is from the library RPC1-11.1 constructed by the group		
	of Pieter de Jong. For further details see		
	http://www.chori.org/bacpac/home.htm		
	VECTOR: pBACe3.6		
FEATURES			
source	Location/Qualifiers		
	1..96183		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="13"		
	/clone="RP11-261P24"		
	/clone_lib="RPC1-11.1"		
repeat_region	1301..1335		
	/note="SN5 repeat: matches 34..68 of consensus"		
repeat_region	1432..1506		
	/note="MER20 repeat: matches 1..75 of consensus"		
repeat_region	3518..3954		
	/note="LIMB8 repeat: matches 5837..6173 of consensus"		
repeat_region	5528..5599		
	/note="12 copies 6 mer tatta 69% conserved"		
repeat_region	5533..5604		



```
repeat_region /note="36 copies 2 mer at 69% conserved"
repeat_region 5766.. .6068
repeat_region /note="AluSp repeat: matches 7. .303 of consensus"
repeat_region 6949.. 7264
repeat_region /note="AluJo repeat: matches 1. .307 of consensus"
repeat_region 7380.. 7501
repeat_region /note="MIR repeat: matches 65. .191 of consensus"
repeat_region 8192.. 8407
repeat_region /note="L2 repeat: matches 2064. .2289 of consensus"
repeat_region 8475.. 8506
repeat_region /note="16 copies 2 mer tt 100% conserved"
repeat_region 8929.. .9103
repeat_region /note="MER33 repeat: matches 4. .171 of consensus"
repeat_region 9104.. .9421
repeat_region /note="AluJb repeat: matches 1. .306 of consensus"
repeat_region 9422.. 9450
repeat_region /note="MER33 repeat: matches 171. .216 of consensus"
repeat_region 9491.. .9595
repeat_region /note="MER33 repeat: matches 142. .225 of consensus"
repeat_region 9596.. .9909
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
repeat_region 9910.. 9997
repeat_region /note="MER33 repeat: matches 225. .319 of consensus"
repeat_region 11665.. .11763
repeat_region /note="MIR repeat: matches 47. .149 of consensus"
repeat_region 11991.. .12022
repeat_region /note="8 copies 4 mer tgtg 90% conserved"
repeat_region 12274.. 12568
repeat_region /note="AluXx repeat: matches 20. .311 of consensus"
repeat_region 12577.. .12879
repeat_region /note="AluY repeat: matches 1. .304 of consensus"
repeat_region 13406.. .13549
repeat_region /note="MLT1B repeat: matches 1. .146 of consensus"
repeat_region 13554.. 13585
repeat_region /note="16 copies 2 mer ca 96% conserved"
repeat_region 13682.. .13715
repeat_region /note="17 copies 2 mer ac 97% conserved"
repeat_region 13684.. .13715
repeat_region /note="8 copies 4 mer acac 100% conserved"
repeat_region 13752.. .13779
repeat_region /note="14 copies 2 mer ca 96% conserved"
repeat_region 13784.. .14048
repeat_region /note="MLT1C repeat: matches 189. .465 of consensus"
repeat_region 14716.. .15018
repeat_region /note="AluSc repeat: matches 1. .303 of consensus"
repeat_region 15567.. .15842
repeat_region /note="AluJo repeat: matches 3. .305 of consensus"
repeat_region 16752.. .16916
repeat_region /note="MIR repeat: matches 63. .227 of consensus"
repeat_region 17700.. .17838
repeat_region /note="AluJo/FRAM repeat: matches 163. .301 of consensus"
repeat_region 18111.. .18314
repeat_region /note="MLT1A2 repeat: matches 157. .359 of consensus"
repeat_region 18317.. .18428
repeat_region /note="28 copies 4 mer ctcc 71% conserved"
repeat_region 18319.. .18422
repeat_region /note="52 copies 2 mer ct 63% conserved"
repeat_region 18322.. .18423
repeat_region /note="17 copies 6 mer tctctc 53% conserved"
repeat_region 18510.. .18797
repeat_region /note="AluJo repeat: matches 1. .302 of consensus"
repeat_region 18800.. .18956
repeat_region /note="MLT1A2 repeat: matches 1. .171 of consensus"
repeat_region 19392.. .19704
repeat_region /note="AluJo repeat: matches 1. .312 of consensus"
repeat_region 19780.. .20069
repeat_region /note="AluY repeat: matches 1. .290 of consensus"
repeat_region 22285.. .22456
repeat_region /note="MIR repeat: matches 23. .220 of consensus"
repeat_region 22457.. .22760
repeat_region /note="AluXx repeat: matches 1. .304 of consensus"
repeat_region 22761.. .22773
repeat_region /note="MIR repeat: matches 220. .230 of consensus"
```

```
repeat_region 26734.. .26961
repeat_region /note="L1MB2 repeat: matches 5944. .6170 of consensus"
repeat_region 28074.. .28174
repeat_region /note="MER44C repeat: matches 7. .103 of consensus"
repeat_region 28175.. .28840
repeat_region /note="MER72 repeat: matches 5. .668 of consensus"
repeat_region 28841.. .29431
repeat_region /note="MER44C repeat: matches 103. .721 of consensus"
repeat_region 29821.. .30105
repeat_region /note="AluXx repeat: matches 1. .285 of consensus"
repeat_region 30219.. .30775
repeat_region /note="L2 repeat: matches 2150. .2749 of consensus"
repeat_region 30924.. .31207
repeat_region /note="AluXx repeat: matches 3. .294 of consensus"
repeat_region 32053.. .32298
repeat_region /note="MIR repeat: matches 1. .257 of consensus"
repeat_region 32498.. .32740
repeat_region /note="L1MB3 repeat: matches 5943. .6182 of consensus"
repeat_region 33867.. .33920
repeat_region /note="27 copies 2 mer ca 90% conserved"
repeat_region 33871.. .33918
repeat_region /note="8 copies 6 mer cacaca 93% conserved"
repeat_region 33878.. .33921
repeat_region /note="11 copies 4 mer acac 97% conserved"
repeat_region 34367.. .34420
repeat_region /note="9 copies 6 mer ctcttc 75% conserved"
repeat_region 34458.. .34576
repeat_region /note="MIR repeat: matches 94. .219 of consensus"
repeat_region 34914.. .35006
repeat_region /note="MIR repeat: matches 150. .242 of consensus"
repeat_region 35044.. .35414
repeat_region /note="MLT1B repeat: matches 1. .386 of consensus"
repeat_region 35440.. .35480
repeat_region /note="L2 repeat: matches 2423. .2464 of consensus"
repeat_region 35653.. .36028
repeat_region /note="THE1B repeat: matches 1. .359 of consensus"
repeat_region 36683.. .37012
repeat_region /note="AluXx repeat: matches 1. .308 of consensus"
repeat_region 37526.. .37818
repeat_region /note="AluJo repeat: matches 1. .290 of consensus"
repeat_region 37957.. .38028
repeat_region /note="36 copies 2 mer tg 86% conserved"
repeat_region 37968.. .38031
repeat_region /note="16 copies 4 mer gtgt 90% conserved"
repeat_region 37980.. .38027
repeat_region /note="8 copies 6 mer gtgtgt 100% conserved"
repeat_region 38553.. .39205
repeat_region /note="L1MB3 repeat: matches 5545. .6183 of consensus"
repeat_region 39561.. .39944
repeat_region /note="MSTA repeat: matches 1. .426 of consensus"
repeat_region 40827.. .41086
repeat_region /note="AluXx repeat: matches 1. .259 of consensus"
repeat_region 41793.. .42089
repeat_region /note="AluXg repeat: matches 1. .297 of consensus"
repeat_region 42330.. .42393
repeat_region /note="L2 repeat: matches 2627. .2701 of consensus"
repeat_region 42692.. .42857
repeat_region /note="L1MC2 repeat: matches 5654. .5816 of consensus"
repeat_region 42858.. .43151
repeat_region /note="AluXq repeat: matches 1. .293 of consensus"
repeat_region 43152.. .43183
repeat_region /note="L1MC2 repeat: matches 5816. .5846 of consensus"
repeat_region 43184.. .43479
```

Query Match 10.8%; Score 254; DB 9; Length 96183;  
Best Local Similarity 100.0%; Pred. No. 3.4e-50;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1060 CAGTCTCAGCAGCAGCAGCAGCCTTACATTTGGAGAGGTGCCGAATCTCCAGGGGCCAG 1119  
|||||  
Db 95665 CAGTCTCAGCAGCAGCAGCAGCCTTACATTTGGAGAGGTGCCGAATCTCCAGGGGCCAG 95724  
Oy 1120 AGCTCCGCGGAGGAGAAAGCAACCGAAGGTTTCCGCCGGGAGCGGGGTGCGACCCGAGC 1179

```
|||||
Db 95725 AGCTCGCGCGAGGAACCGAAGTTTCGCGGGGAGCGGGTCGCACCCGAGC 95784
|||||
QY 1180 CTTGCGCGCGAGAGAGCCCGCGGGTAAACAGAGCGGAGCGGAGCGCTCGCGCGCC 1239
|||||
Db 95785 CTTGCGCGCGAGAGAGCCCGCGGGTAAACAGAGCGGAGCGGAGCGCTCGCGCGCC 95844
|||||
QY 1240 ACGGAGGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGG 1299
|||||
Db 95845 ACGGAGGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGG 95904
|||||
QY 1300 CAGCCAAGCACAGG 1313
|||||
Db 95905 CAGCCAAGCACAGG 95918
|||||

RESULT 10
AL137249 Human DNA sequence from clone RP11-111L24 on chromosome
DEFINITION 13q31.3-32.3, complete sequence.
ACCESSION AL137249
VERSION AL137249.29 GI:18655955
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 106578)
Garner,P.
Direct Submission
Submitted (12-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humqueresanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 13, 2002 this sequence version replaced gi:18121451.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-111L24 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-111L24. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-295B17 is at 104579 in this
sequence. The true right end of clone RP11-261P24 is at 2000 in
this sequence.
Location/Qualifiers
1. 106578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q31.3-32.3"

FEATURES
source
```

```
misc_feature
/clone="RP11-111L24"
/clone_lib="RPCI-11.1"
79533
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 800bp by restriction digest data"
BASE COUNT 28669 a 25979 c 25140 g 26790 t
ORIGIN

Query Match 10.8% Score 254; DB 9; Length 106578;
Best Local Similarity 100.0%; Pred. No. 3.5e-50;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1060 CAGTCTCAGCAGAGCACCAGCCTTACATTGAGAGAGTGCCGAATCTCCAGGGGGCCAG 1119
|||||
Db 1482 CAGTCTCAGCAGAGCACCAGCCTTACATTGAGAGAGTGCCGAATCTCCAGGGGGCCAG 1541
|||||
QY 1120 AGCTCGCGCGAGGAAAGAACCGAAGTTTCGCGGGGAGCGGGTCGCACCCGAGC 1179
|||||
Db 1542 AGCTCGCGCGAGGAAAGAACCGAAGTTTCGCGGGGAGCGGGTCGCACCCGAGC 1601
|||||
QY 1180 CTTGCGCGCGAGAGAGCCCGCGGGTAAACAGAGCGGAGCGGAGCGCTCGCGCGCC 1239
|||||
Db 1602 CTTGCGCGCGAGAGAGCCCGCGGGTAAACAGAGCGGAGCGGAGCGCTCGCGCGCC 1661
|||||
QY 1240 ACGGAGGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGG 1299
|||||
Db 1662 ACGGAGGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGG 1721
|||||
QY 1300 CAGCCAAGCACAGG 1313
|||||
Db 1722 CAGCCAAGCACAGG 1735
|||||

RESULT 11
AC022669/c AC022669 152053 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 13 clone RP11-96B23 map 13, WORKING DRAFT
DEFINITION SEQUENCE, 33 unordered pieces.
ACCESSION AC022669
VERSION AC022669.4 GI:9966291
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152053)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
AUTHORS
REFERENCE 1 (bases 1 to 152053)
TITLE Homo sapiens chromosome 13, clone RP11-96B23
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 152053)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczký,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 152053)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
```

Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collimore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glnde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenca, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 4, 2000 this sequence version replaced gi:7229803.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center -----  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIDR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information -----  
Center project name: L4938  
Center clone name: 96\_B\_23  
----- Summary Statistics -----  
Sequencing vector: M13; M77815; 93% of reads  
Sequencing vector: Plasmid; n/a; %0.f% of reads  
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 136451 bases at least Q40  
Consensus quality: 143612 bases at least Q30  
Consensus quality: 146365 bases at least Q20  
Insert size: 158000; agarose-fp  
Insert size: 148853; sum-of-contigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality cov.  
NOTE: this is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 784: contig of 784 bp in length  
785 884: gap of 100 bp  
885 1956: contig of 1072 bp in length  
1957 2056: gap of 100 bp  
2057 3113: contig of 1057 bp in length  
3114 3213: gap of 100 bp  
3214 4817: contig of 1604 bp in length  
4818 4917: gap of 100 bp  
4918 6582: contig of 1665 bp in length  
6583 6682: gap of 100 bp  
6683 8059: contig of 1377 bp in length  
8060 8159: gap of 100 bp  
8160 22369: contig of 14210 bp in length  
22370 22469: gap of 100 bp  
22470 24974: contig of 2505 bp in length  
24975 25074: gap of 100 bp  
25075 27206: contig of 2132 bp in length  
27207 27306: gap of 100 bp

27307 29420: contig of 2114 bp in length  
29421 29520: gap of 100 bp  
29521 32000: contig of 2480 bp in length  
32001 32100: gap of 100 bp  
32101 36330: contig of 4230 bp in length  
36331 36430: gap of 100 bp  
36431 39067: contig of 2637 bp in length  
39068 39167: gap of 100 bp  
39168 42564: contig of 3397 bp in length  
42565 42664: gap of 100 bp  
42665 46664: contig of 4000 bp in length  
46665 46764: gap of 100 bp  
46765 51047: contig of 4283 bp in length  
51048 51147: gap of 100 bp  
51148 55269: contig of 4122 bp in length  
55270 55369: gap of 100 bp  
55370 58831: contig of 3462 bp in length  
58832 58931: gap of 100 bp  
58932 64077: contig of 5146 bp in length  
64078 64177: gap of 100 bp  
64178 68808: contig of 4631 bp in length  
68809 68908: gap of 100 bp  
68909 74170: contig of 5262 bp in length  
74171 74270: gap of 100 bp  
74271 79296: contig of 5026 bp in length  
79297 79396: gap of 100 bp  
79397 84724: contig of 5328 bp in length  
84725 84824: gap of 100 bp  
84825 89713: contig of 4889 bp in length  
89714 89813: gap of 100 bp  
89814 94721: contig of 4908 bp in length  
94722 94821: gap of 100 bp  
94822 100196: contig of 5375 bp in length  
100197 100296: gap of 100 bp  
100297 107015: contig of 6719 bp in length  
107016 107115: gap of 100 bp  
107116 114507: contig of 7392 bp in length  
114508 114607: gap of 100 bp  
114608 122671: contig of 8064 bp in length  
122672 122771: gap of 100 bp  
122772 128052: contig of 5281 bp in length  
128053 128152: gap of 100 bp  
128153 138583: contig of 10431 bp in length  
138584 138683: gap of 100 bp  
138684 150248: contig of 11565 bp in length  
150249 150348: gap of 100 bp  
150349 152053: contig of 1705 bp in length.

FEATURES  
Location/Qualifiers  
1..152053  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13"  
/clone="RP11-96B23"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..784  
/note="assembly\_fragment"  
vector\_end:SP6  
vector\_side:left  
885..1956  
/note="assembly\_fragment"  
2057..3113  
/note="assembly\_fragment"  
3214..4817  
/note="assembly\_fragment"  
4918..6582  
/note="assembly\_fragment"  
6683..8059  
/note="assembly\_fragment"  
8160..22369  
/note="assembly\_fragment"  
22470..24974  
/note="assembly\_fragment"

misc\_feature  
1  
784: contig of 784 bp in length  
785 884: gap of 100 bp  
885 1956: contig of 1072 bp in length  
1957 2056: gap of 100 bp  
2057 3113: contig of 1057 bp in length  
3114 3213: gap of 100 bp  
3214 4817: contig of 1604 bp in length  
4818 4917: gap of 100 bp  
4918 6582: contig of 1665 bp in length  
6583 6682: gap of 100 bp  
6683 8059: contig of 1377 bp in length  
8060 8159: gap of 100 bp  
8160 22369: contig of 14210 bp in length  
22370 22469: gap of 100 bp  
22470 24974: contig of 2505 bp in length  
24975 25074: gap of 100 bp  
25075 27206: contig of 2132 bp in length  
27207 27306: gap of 100 bp

misc\_feature  
885..1956  
/note="assembly\_fragment"  
2057..3113  
/note="assembly\_fragment"  
3214..4817  
/note="assembly\_fragment"  
4918..6582  
/note="assembly\_fragment"  
6683..8059  
/note="assembly\_fragment"  
8160..22369  
/note="assembly\_fragment"  
22470..24974  
/note="assembly\_fragment"

TITLE  
JOURNAL

## COMMENT









DNANAVKSIPIKDVPIVHETKITYEAAQTDDNSDLDPGVLLTAQTIITSETPSST  
TTTQITKTIVKGGISITREKIRIVITGDADIDHDQVLVQAIKEAKEQHPDMSVTKVVVH  
QTEIADE\*

Search completed: December 3, 2002, 20:46:10  
Job time : 5413.44 secs

BASE COUNT 881 a 563 c 601 g 550 t  
ORIGIN

Query Match 8.3%; Score 194; DB 9; Length 2595;

Best Local Similarity 52.9%; Pred. No. 8e-36;

Matches 491; Conservative 0; Mismatches 425; Indels 12; Gaps 3;

QY	15	CGTGTCCATCCAGATGCTGGATGACACCCAGGAGGATTTGAAGTTCCACAAAG	74
Db	627	CATGCACTGCAAGTTTCTTTGGTGTGACACAGTTTATGAATGTTGTGGAGAAACA	686
QY	75	AGCTCCTGGGAAGTCTGCTGGATGACAGTTTGCACACCTCAACCTCGTGGAGGTGA	134
Db	687	TGCTAAGGGACAGATTGCTTAAACGAGTATGTGAGCATCTCAATCTTTGGAGAAGA	746
QY	135	CTATTTTGGCTCGAGTTTCTGTATCACAAGAAAGATCAGGGTGTGGCTGGATCTCTAA	194
Db	747	CTATTTTGGCTAGCCATTGGGATAACACACCTCTAAGACATGGCTGGATTCGCGCAA	806
QY	195	ACCATGTGTGAACAGATTAGAGGCCAAAGCAGCTGTGTGTTAACTTTGTGCTGAAAT	254
Db	807	AGNAATAAAAAGCAGGTTCGGG---TGTCCCTTGGAAATTTTACATTTAATGTAAGTT	863
QY	255	CTTTCGCTGACACACACAACTCCAAAGAAAGAACTCACAAGGTACCTGTTCGCGTGCA	314
Db	864	TTATCCACCTGACCCAGCAGTTAACAGAAGACATAACAAGATATATTATGCTCTCA	923
QY	315	GGTGAAGCAGGACTTGGCTCAGGCAAGTTGAGGTGTAATGACACAGCGCAGCTCTCT	374
Db	924	GCCTTCGGCAGGACATAGTTGCAGGAGCTGTGCCCTGTCTCTTGCACACCTTAGCAT	983
QY	375	GATTTCCACATTTGCTCAATCTGAGATTTGGGATTTTGTGATGAGCCTTGGACAGAGCA	434
Db	984	AGTTCTTACACCATCCAGCTGGAAGTGGGAGACTAGACCCAGAACTCCATGGCGTGA	1043
QY	435	CTTAGCAAAAATAAATACATACC-----TCAGCAAGACGCACTAGAGCACAATAATCGT	488
Db	1044	TTATGTTAGTGATTTTAAACTGGCCCCGAATCAGACCAAGGAACCTGAAGAGAAGTCAT	1103
QY	489	GGNATTTACACATAACCAATTGGGCAAAACACACAGCAGAAATCAGATTTCCAGCTCTAGA	548
Db	1104	GGAACTGCATAGTCATACAGGTCCATGACTCCAGCTCAGGCTGACTTGGAGTTCTTGA	1163
QY	549	GATTGCCGTGCGGTAGAGATGATGGAATCCGGTTGCACCCGCCCAAGGACAGGGAAG	608
Db	1164	GAATGCCAAAAGTTGCTATGATGGAGTTGATCTTCATAAGCAAGGACTTGGGAAG	1223
QY	609	CACGAAGATCAATCTGGCGGTTGCCAACACGGAATTTAGTGTTCAGGGTTTCACTAA	668
Db	1224	AGTAGATATCATCTAGGTGTCTGTCTAGTGGCTTCTGGTTTACAAGATAAGCTGAG	1283
QY	669	GATCAATGCTTCACTGGGCCAAGGTGGGGAAGCTGAGCTTCAAGAGGAAGCGTTTCT	728
Db	1284	AATTAACCGGTTCCCTTGGCCCCAAAGTGTGAAGATTTCTTATAAACCTAGTAGCTTTT	1343
QY	729	CATCAAGCTCCGGCAGATGCCAATAGTGGTACACAGATACCTTGGAAATTCCTGATGC	788
Db	1344	CATCAAGATTCGGCCCTGGAGCAAGAGCAGTATGAAGTACCATCCGATTCAAACTTCC	1403
QY	789	CAGTCGGGATTTCTGCAAGTCTCTCTGAAAATCTGTGTGAACATCATGCCCTTCTTTAG	848
Db	1404	CAGTTACCGAGCAGCTAAGAAATATGGAAGTCTGTGTAGAACATCACACGTTTTTCAG	1463
QY	849	ACTTTTGAAGCCCCAACCAAGCCCGCTCTTTTAGCCGGGGGTTCATCAT	908
Db	1464	A---TTGACATCTACAGACACCATTCCTCCAAAAGCAAAATTTCTTGGCTAGGATCCAAAT	1520
QY	909	TCGGTTCAGTGTGGGACTCAGAAGCAG	936
Db	1521	TCGATACAGTGGCGGACTCAAGCTCAG	1548



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: December 3, 2002, 14:48:04 ; Search time 371.636 Seconds  
(without alignments)  
14246.337 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_151\_2501

Perfect score: 2351

Sequence: 1 ccttcaggaactcgtgtc.....attgaggagcgagacga 2351

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2351	100.0	3442	AA79183	Human chondrocyte-
2	2349.4	99.9	4687	21 AAA08582	Human cytoskeleton
3	2335.2	99.3	3187	23 AAS64760	DNA encoding novel
c	5	254	10.8	1824 23 AAC98992	Human pancreatic c
	6	228	9.7	1220 23 AAS64761	DNA encoding novel
c	7	223.6	9.5	240 23 ABL02335	Drosophila melanog
	8	192.4	8.2	4922 21 AAA42150	Human secreted exp
9	181.8	7.7	3643	24 ABV30214	Human prostate exp
				24 ABK84074	Human cDNA differe

10	180.4	7.7	6263	24	ABK70019	Human cDNA for the
11	171.4	7.3	2916	23	ABL12469	Drosophila melanog
12	171.4	7.3	3470	23	ABL18111	Drosophila melanog
13	171.4	7.3	5897	23	ABL18211	Drosophila melanog
14	169.6	7.2	3620	21	AAA08581	Human cytoskeleton
15	162.6	6.9	14917	24	AAD32338	Human lung specifi
16	154.6	6.6	426	23	AAS58814	cDNA #1490 encodin
17	144	6.1	3984	18	AAT58627	Protein tyrosine p
18	144	6.1	3984	20	AAK78463	Human PTPH1 cDNA.
19	144	6.1	3984	24	AAI72918	PTPH1 cDNA. Homo
20	142.8	6.1	2872	22	AAS08887	Human protein tyro
21	142.8	6.1	3113	23	ABN93400	Human gene GS91552
22	138.2	5.9	3106	23	ABN93399	Human gene GS91552
23	134.2	5.7	3031	23	ABV22627	Human prostate exp
24	134.2	5.7	3031	23	ABV28450	Human prostate exp
25	133.6	5.7	2067	24	ABN85314	Human cytoskeleton
26	133.6	5.7	3166	22	AAH17787	Human cDNA sequenc
27	133.4	5.7	4356	23	ABL29969	Drosophila melanog
28	132.2	5.6	2846	24	ABL60869	Human cytoskeleton
29	121.8	5.2	2337	24	ABQ73378	Human red blood ce
30	106	4.5	1718	22	AAS27016	cDNA encoding nove
31	105.8	4.5	1747	23	ABL03867	Drosophila melanog
32	105.8	4.5	4150	23	ABL03866	Drosophila melanog
33	105.2	4.5	716	22	AAS27440	cDNA encoding nove
34	104.6	4.4	7982	23	ABL02334	Drosophila melanog
35	102.8	4.4	3162	24	ABQ72637	Human MDDT encodin
36	102.4	4.4	6411	23	ABL02196	Drosophila melanog
37	102.2	4.3	621	24	ABQ72602	Human MDDT encodin
38	99.4	4.2	531	22	AAI22865	Human breast cance
39	98.4	4.2	2066	24	ABL60867	Human cytoskeleton
40	98	4.2	717	24	ABL90828	Human polynucleoti
41	95.8	4.1	659	22	AAH33845	Human colon cancer
42	90.2	3.8	544	22	AAH98023	Murine 7-transmemb
43	87	3.7	3260	23	ABL02197	Drosophila melanog
44	83.6	3.6	665	22	AAH07867	Human cDNA clone (
45	83.6	3.6	1296	24	ABK49898	Human cDNA encodin

ALIGNMENTS

RESULT 1  
AA79183  
ID AA79183 standard; DNA; 3442 BP.

XX AA79183;

AC AA79183;

XX 17-AUG-1999 (first entry)

XX Human chondrocyte-derived gene CDEP.

DE Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;

XX Dbl homology domain; pleckstrin homology domain; rheumatoid 'arthritis;

XX drug; ss.

XX Homo sapiens.

XX WO9928458-A1.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-JP05348.

XX 27-NOV-1997; 97JP-0342060.

XX (CHUS ) CHUGAI SEIYAKU KK.

XX Kato Y, Kawamoto T, Koyano Y;

XX WPI; 1999-371117/31.

XX P-PSDB; AAY07482.

XX Protein CDEP expressed in differentiated chondrocytes, and gene

PT encoding it  
XX PS Claim 5; Fig 1; 59pp; Japanese.  
XX  
CC This sequence represents the coding region for a protein (CDEP) expressed  
CC in differentiated human foetal chondrocytes, which contains an ezrin-like  
CC domain, a Dbl homology (DH) domain and a pleckstrin homology (PH) domain.  
CC The nucleic acid or protein can be used in the investigating and  
CC treatment of cancers and arthritic diseases (including chronic rheumatoid  
CC arthritis), or for screening of candidate anticancer drugs.  
XX  
SQ Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;  
Query Match 100.0%; Score 2351; DB 20; Length 3442;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 COTTGAGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCAATT 60  
DB 151 CTTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCAATT 210  
QY 61 GAAGTCCACAAGAGCTCTCGGAGAGTGTCTGGATGAGTTGCAACCACTCAAC 120  
DB 211 GAAGTCCACAAGAGCTCTCGGAGAGTGTCTGGATGAGTTGCAACCACTCAAC 270  
QY 121 CTCGTGGAAGTGACTATTTTGGCCCTCGAGTTTCTGATCACAAAAGATCACGGTGTGG 180  
DB 271 CTCGTGGAAGTGACTATTTTGGCCCTCGAGTTTCTGATCACAAAAGATCACGGTGTGG 330  
QY 181 CTGGATCTCTAAACCACTTGTGAACAGATTAGAAGGCCAAAGCAGCTGTGTTTAAG 240  
DB 331 CTGGATCTCTAAACCACTTGTGAACAGATTAGAAGGCCAAAGCAGCTGTGTTTAAG 390  
QY 241 TTTGTGGTGAATCTTTCGCTCACCACACACAACTCCAGAGAACTCACAGGTAC 300  
DB 391 TTTGTGGTGAATCTTTCGCTCACCACACAACTCCAGAGAACTCACAGGTAC 450  
QY 301 CTGTTCCGCTGCAGGTGAAGCAGACTTGGCTCAAGCAGGTTCAGCTGTAATGACACC 360  
DB 451 CTGTTCCGCTGCAGGTGAAGCAGACTTGGCTCAAGCAGGTTCAGCTGTAATGACACC 510  
QY 361 AGCGAGCTCTCTGATTTACACATTTGTGCAATCTGAGATTTGGGATTTGATGAAGCC 420  
DB 511 AGCGAGCTCTCTGATTTACACATTTGTGCAATCTGAGATTTGGGATTTGATGAAGCC 570  
QY 421 TTGGACAGAGAGCACTTAGCAAAAATAAATACATACCTCAGCAAGACGCATAGAGAC 480  
DB 571 TTGGACAGAGAGCACTTAGCAAAAATAAATACATACCTCAGCAAGACGCATAGAGAC 630  
QY 481 AAAATCGTGGAAATTCACCAATAAACCACATTTGGACAAACACCCAGCAGATTCACG 540  
DB 631 AAAATCGTGGAAATTCACCAATAAACCACATTTGGACAAACACCCAGCAGATTCACG 690  
QY 541 CTCCTAGAGATTTGCCGCTGAGATGATGAAATCCCGTTGACCGGCCCAAGAC 600  
DB 691 CTCCTAGAGATTTGCCGCTGAGATGATGAAATCCCGTTGACCGGCCCAAGAC 750  
QY 601 AGGGAAGCAGCAAGATCAATCTGCGCTTGCACACCGGAAATCTAGTGTTCAGGGT 660  
DB 751 AGGGAAGCAGCAAGATCAATCTGCGCTTGCACACCGGAAATCTAGTGTTCAGGGT 810  
QY 661 TTCCTAAGATCAATGCTTCAACTGGGCCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAG 720  
DB 811 TTCCTAAGATCAATGCTTCAACTGGGCCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAG 870  
QY 721 CGCTTTCATCAAGCTCCGGCAGATGCCAATAGTCGTACCAAGGATACCTTGGAAATTC 780  
DB 871 CGCTTTCATCAAGCTCCGGCAGATGCCAATAGTCGTACCAAGGATACCTTGGAAATTC 930  
QY 781 CTGATGGCCAGTCGGATTTCTGCAAGTCTTCTGGAATACTGTGTTGAACATCATGCC 840  
DB 931 CTGATGGCCAGTCGGATTTCTGCAAGTCTTCTGGAATACTGTGTTGAACATCATGCC 990

QY 841 TTCTTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCCTCTCTTTAGCGGGG 900  
DB 991 TTCTTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCCTCTCTTTAGCGGGG 1050  
QY 901 TCATCATTTTCGGTTCAGTGTGCGACTCAGAAGCAGGTTCGACTATGTTAAAGAAGGA 960  
DB 1051 TCATCATTTTCGGTTCAGTGTGCGACTCAGAAGCAGGTTCGACTATGTTAAAGAAGGA 1110  
QY 961 GGACATAGAAGGTGACGTTTGAAGGAAGCAGACAGATTCATTCTATCCGAGGCCCT 1020  
DB 1111 GGACATAGAAGGTGACGTTTGAAGGAAGCAGACAGATTCATTCTATCCGAGGCCCT 1170  
QY 1021 GCTTCACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCTCAGCAGACGACCCAGC 1080  
DB 1171 GCTTCACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCTCAGCAGACGACCCAGC 1230  
QY 1081 CTTTACATTTGGAAGAGTGCAGAACTCTCCAGGGGCCAGAGCTGCCGCGGAGGAAGGAA 1140  
DB 1231 CTTTACATTTGGAAGAGTGCAGAACTCTCCAGGGGCCAGAGCTGCCGCGGAGGAAGGAA 1290  
QY 1141 CCGAGGTTTCCGCGGGAGCGGGGTGCACCCGAGCCCTGCGCCGAGAGAGCCCC 1200  
DB 1291 CCGAGGTTTCCGCGGGAGCGGGGTGCACCCGAGCCCTGCGCCGAGAGAGCCCC 1350  
QY 1201 CCGGTAAACAGCAGCGGAGCGCCCTCGGCGCCACGAGAGAGAGAGAGGTC 1260  
DB 1351 CCGGTAAACAGCAGCGGAGCGCCCTCGGCGCCACGAGAGAGAGAGGTC 1410  
QY 1261 GTTAAAGATAGAGCCAGCAGAGTAACCTCAGCCCCCGCAGCCAAAGCAGAGTCCCTG 1320  
DB 1411 GTTAAAGATAGAGCCAGCAGAGTAACCTCAGCCCCCGCAGCCAAAGCAGAGTCCCTG 1470  
QY 1321 ACTGSCAGTCTCAGCTTTCGAGCTGTCTGTAACCTCGAGGGGGAGTGGCCCTGCC 1380  
DB 1471 ACTGSCAGTCTCAGCTTTCGAGCTGTCTGTAACCTCGAGGGGGAGTGGCCCTGCC 1530  
QY 1381 AACGTGACCTTGTCTCCCAACTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAGC 1440  
DB 1531 AACGTGACCTTGTCTCCCAACTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAGC 1590  
QY 1441 CCGTGTGTAATGACAGGCTGCCCGGAGCGAGCATGAGGATGAGGGCGGAGGAAG 1500  
DB 1591 CCGTGTGTAATGACAGGCTGCCCGGAGCGAGCATGAGGATGAGGGCGGAGGAAG 1650  
QY 1501 AGATTCCCACTGATAAGCGTACTTCTAGTAAGGAAGTGTCTACCAACGAGGGAACA 1560  
DB 1651 AGATTCCCACTGATAAGCGTACTTCTAGTAAGGAAGTGTCTACCAACGAGGGAACA 1710  
QY 1561 TATCTGAAGGATCTCGAAGTTATCATTCTCGTGTTCAGACACAGTGAGCAAAAGAGAC 1620  
DB 1711 TATCTGAAGGATCTCGAAGTTATCATTCTCGTGTTCAGACACAGTGAGCAAAAGAGAC 1770  
QY 1621 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGGAATTTTGAACCTTTGCACAAATTT 1680  
DB 1771 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGGAATTTTGAACCTTTGCACAAATTT 1830  
QY 1681 CATCTAATTTTCTCAAGGAAATTCAGCAACGACTTGGCCCTGTGGGAAGCGCGCTCAAT 1740  
DB 1831 CATCTAATTTTCTCAAGGAAATTCAGCAACGACTTGGCCCTGTGGGAAGCGCGCTCAAT 1890  
QY 1741 GCCCAATACAGAGATTACCAAGAAATCGGCGATGTCTGCTGAAGAACAATTCAGGGCATG 1800  
DB 1891 GCCCAATACAGAGATTACCAAGAAATCGGCGATGTCTGCTGAAGAACAATTCAGGGCATG 1950  
QY 1801 AAGCACCTGGGGCTCAGCTGTGGAAGCAGACGAGGCCCTTGGAGGCCCTGGAGAAATGGA 1860  
DB 1951 AAGCACCTGGGGCTCAGCTGTGGAAGCAGACGAGGCCCTTGGAGGCCCTGGAGAAATGGA 2010  
QY 1861 ATCAAGAGCTCCCGCGGCTGGAAACTTCTGCAGAGACTTTGAGCTCAGAAAGTGTGT 1920  
DB 2011 ATCAAGAGCTCCCGCGGCTGGAAACTTCTGCAGAGACTTTGAGCTCAGAAAGTGTGT 2070  
QY 1921 TACCTACGCTCAACACCTTCTCCTGCGGCCACTGACCGGCTCATGCAAGCAG 1980

```
Db 2071 TACCTACCGCTCAACACTTCTCCTCGCGCCACTGACCGGCTCATGCACTACAAGCAG 2130
QY 1981 GTCTCGAGCGGTGTGCAAAACACCCCGCGAGCCACGCCGACTTCAGGGACTGCCGA 2040
Db 2131 GTCTCGAGCGGTGTGCAAAACACCCCGCGAGCCACGCCGACTTCAGGGACTGCCGA 2190
QY 2041 GCCGCTTTGGCAGAGATCACGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATG 2100
Db 2191 GCCGCTTTGGCAGAGATCACGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATG 2250
QY 2101 GAGAAATTCAGAGCTGCACGAACTCAAGAAAGATTGGATTGGCATTGACATCTTGTG 2160
Db 2251 GAGAAATTCAGAGCTGCACGAACTCAAGAAAGATTGGATTGGCATTGACATCTTGTG 2310
QY 2161 GTTCCGGGAAGGAGTTCTATCCCTCTGGGAGCTCAGCAAGCTCTCGGGGAAGGGGCTC 2220
Db 2311 GTTCCGGGAAGGAGTTCTATCCCTCTGGGAGCTCAGCAAGCTCTCGGGGAAGGGGCTC 2370
QY 2221 CAGCAGGCGATGTTCTTCTGTTCAAGAGCTCCTGCTATACAGAGCCGGGGCTGACG 2280
Db 2371 CAGCAGGCGATGTTCTTCTGTTCAAGAGCTCCTGCTATACAGAGCCGGGGCTGACG 2430
QY 2281 GCCTCAATCAGTTTAAAGTCCAGGGCAGCTCCCGCTCTATGGCATGAGGAG 2340
Db 2431 GCCTCAATCAGTTTAAAGTCCAGGGCAGCTCCCGCTCTATGGCATGAGGAG 2490
QY 2341 AGCGAAGACGA 2351
Db 2491 AGCGAAGACGA 2501

RESULT 2
AAA08582
ID AAA08582 standard; DNA; 4687 BP.
XX
AC AAA08582;
XX
DT 19-JUL-2000 (first entry)
XX
DE Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.
XX
KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;
KW autoimmunity; inflammatory; vesicle trafficking; neurological;
KW cardiovascular; cell motility; reproductive; muscle disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 90..3227
FT /tag= a
FT /product= CYSKP-2
XX
XX WO200017355-A2.
XX
XX 30-MAR-2000.
XX
XX 17-SEP-1999; 99WO-US21565.
XX
XX 18-SEP-1998; 98US-0172226.
XX
XX 27-APR-1999; 99US-0131321.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;
XX Guegler KJ, Patterson C, Azimzai Y, Baughn MR;
XX WPI; 2000-283582/24.
XX P-PSDB; AAY91947.
XX
XX Human cytoskeleton associated proteins, used to treat cell
XX proliferative, autoimmune/inflammatory, vesicle trafficking,
XX neurological, cell motility, reproductive and muscle disorders
```

```
xx
PS Claim 9; Page 101-102; 113pp; English.
xx
CC AAA08581-96 encode human cytoskeleton associated proteins 1 to 16
CC (CYSKP-1 to CYSKP-16) respectively. The sequences can be used to treat
CC and diagnose cancer and cell proliferative, autoimmune/inflammatory,
CC vesicle trafficking, neurological, cardiovascular, cell motility,
CC reproductive and muscle disorders. Pharmaceutical compositions
CC containing CYSKP-2 to CYSKP-16 can be used to treat or prevent disorders
CC associated with decreased expression or activity of CYSKP (claimed), for
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,
CC cancers, autoimmune/antimflammatory disorders such as allergies, anemia,
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
CC and trauma. CYSKP antagonists can be used to treat or prevent a
CC disorder associated with increased expression or activity of CYSKP
CC (claimed).
xx
SQ Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;

Query Match 99.9%; Score 2349.4; DB 21; Length 4687;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTCTCAGGAAACTCGTGTCATCAAAATCCAGATCGTGGATGACACCCAGGAGCATTT 60
Db 192 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATCGTGGATGACACCCAGGAGCATTT 251
QY 61 GAAGTTCCACAAAGAGCTCCTGGGAAGGTGCTGCTGGATGCGAGTTTGCACACCTCAAC 120
Db 252 GAAGTTCCACAAAGAGCTCCTGGGAAGGTGCTGCTGGATGCGAGTTTGCACACCTCAAC 311
QY 121 CTGCTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTGG 180
Db 312 CTGCTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTGG 371
QY 181 CTGGATCTCTAAACCCCATTTGAAACAGATTAGAGGCCAAAGCAGCTGTGTGTTAAG 240
Db 372 CTGGATCTCTAAACCCCATTTGAAACAGATTAGAGGCCAAAGCAGCTGTGTGTTAAG 431
QY 241 TTTGTGCTGAAATTTCTCCGCTGACACACACACTCCAAGAAGACTCACAAAGGTAC 300
Db 432 TTTGTGCTGAAATTTCTCCGCTGACACACACACTCCAAGAAGACTCACAAAGGTAC 491
QY 301 CTGTTCCGCTGTCAGGTGAAGCAGAGCTTGGCTCAAGGCAGGTGACGTGTAATGACACC 360
Db 492 CTGTTCCGCTGTCAGGTGAAGCAGAGCTTGGCTCAAGGCAGGTGACGTGTAATGACACC 551
QY 361 AGCGAGCTCTCTTGATTTTACACATTTGCAATCTGAGATTGGGATTTTGATGAAGCC 420
Db 552 AGCGAGCTCTCTTGATTTTACACATTTGCAATCTGAGATTGGGATTTTGATGAAGCC 611
QY 421 TTGCACAGAGAGCATTAGCAAAAATAATATACATACCTCAGCAAGCAGCTAGAGGAC 480
Db 612 TTGCACAGAGAGCATTAGCAAAAATAATATACATACCTCAGCAAGCAGCTAGAGGAC 671
QY 481 AAAATCGTGAATTTTCAACCATTAACCATTTGGACAAACACACAGCAATCAGATTTCCAG 540
Db 672 AAAATCGTGAATTTTCAACCATTAACCATTTGGACAAACACACAGCAATCAGATTTCCAG 731
QY 541 CTCCTAGAGATTGCCGCTGCGGTAGAGATGTATGGAATCCGGTTGCCCGGCAAGGAC 600
Db 732 CTCCTAGAGATTGCCGCTGCGGTAGAGATGTATGGAATCCGGTTGCCCGGCAAGGAC 791
QY 601 AGGGAAGGCACGAAGATCAATCTGGCCGTTGCCAACACGCGAATTTCTAGTGTTCAGGGT 660
Db 792 AGGGAAGGCACGAAGATCAATCTGGCCGTTGCCAACACGCGAATTTCTAGTGTTCAGGGT 851
QY 661 TTCACCTAAGATCAATGCCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAG 720
Db 852 TTCACCTAAGATCAATGCCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAG 911
QY 721 CGCTTTCTCATCAAGCTCCGGCCAGATGCCAATAGTGCCTACCAAGGATACCTTGGAAATTC 780
```

Db	912	CGCTTTTTCATCAAGCTTCGGCCAGATGCCAATAGTCGTACCAAGGATACCTTGGAAATTC	971
Qy	781	CTGATGCCAGTCGGGATTCTCGAAGTCCTTCTCGGAAATCTGTGTGAACATCATGCC	840
Db	972	CTGATGCCAGTCGGGATTCTCGAAGTCCTTCTCGGAAATCTGTGTGAACATCATGCC	1031
Qy	841	TTCTTTAGACTTTTGAAGAGCCAAACAAAGCCCAAGCCGCTCTCTTTAGCCGGGG	900
Db	1032	TTCTTTAGACTTTTGAAGAGCCAAACAAAGCCCAAGCCGCTCTCTTTAGCCGGGG	1091
Qy	901	TCATCATTTCCGTTTCAGTGGTTCGGACTCAGAAGCAGTTCTCGACTATGTTAAGAAGGA	960
Db	1092	TCATCATTTCCGTTTCAGTGGTTCGGACTCAGAAGCAGTTCTCGACTATGTTAAGAAGGA	1151
Qy	961	GGACATAAAGAGTGCAGTTTGAAGAGGAACACAAAGATTCAATCTATCCGGAGCCCT	1020
Db	1152	GGACATAAAGAGTGCAGTTTGAAGAGGAACACAAAGATTCAATCTATCCGGAGCCCT	1211
Qy	1021	GCCTACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCTCAGCAGAGCACCAGC	1080
Db	1212	GCCTACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCTCAGCAGAGCACCAGC	1271
Qy	1081	CTTACATTTCCGAAAGTCCGGAATCTCCAGGGGCCAGAGCTGCCGCCGAGGAAGGAA	1140
Db	1272	CTTACATTTCCGAAAGTCCGGAATCTCCAGGGGCCAGAGCTGCCGCCGAGGAAGGAA	1331
Qy	1141	CCGAAGCTTTCCGCCGGGAGCCGGGTCGCACCCAGCCCTCGCCGCGAGGAGAAGCCCC	1200
Db	1332	CCGAAGCTTTCCGCCGGGAGCCGGGTCGCACCCAGCCCTCGCCGCGAGGAGAAGCCCC	1391
Qy	1201	CGGGTTAAACAGCAGCGGACGGAGCCGCTCGGGCCCCAGGAGGAGAGGAGGTC	1260
Db	1392	CGGGTTAAACAGCAGCGGACGGAGCCGCTCGGGCCCCAGGAGGAGAGGAGGTC	1451
Qy	1261	GTTAAGGATAGGACCCAGCAGAGTAAACTCAGCCCCCGCAGCCAAAGCAGAGCTCCCTG	1320
Db	1452	GTTAAGGATAGGACCCAGCAGAGTAAACTCAGCCCCCGCAGCCAAAGCAGAGCTCCCTG	1511
Qy	1321	ACTGGCAGTCTCACCCTTTCGAGGTGTCGTGTGAACCTCGCAGGGGGAGTGGCCCTGCC	1380
Db	1512	ACTGGCAGTCTCACCCTTTCGAGGTGTCGTGTGAACCTCGCAGGGGGAGTGGCCCTGCC	1571
Qy	1381	AACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAGC	1440
Db	1572	AACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAGC	1631
Qy	1441	CCGCTGCTGAATCACCAGGCTTCCGCCGGAGGAGATGAGGATGAGGCGCCGAGGAAG	1500
Db	1632	CCGCTGCTGAATCACCAGGCTTCCGCCGGAGGAGATGAGGATGAGGCGCCGAGGAAG	1691
Qy	1501	AGATTCCCACTCATAAAGCGTACTTCATAGCTAAGGAAGTGTCTACACCCGAGCGAACA	1560
Db	1692	AGATTCCCACTCATAAAGCGTACTTCATAGCTAAGGAAGTGTCTACACCCGAGCGAACA	1751
Qy	1561	TATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCAGTGACGAAGAGGAC	1620
Db	1752	TATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCAGTGACGAAGAGGAC	1811
Qy	1621	GCCATCCCGGAAGCACTGAAAAGTCTCATATTTCCGAAATTTTGAAACCTTTGCACAAATTT	1680
Db	1812	GCCATCCCGGAAGCACTGAAAAGTCTCATATTTCCGAAATTTTGAAACCTTTGCACAAATTT	1871
Qy	1681	CATCACTAATTTTCTCAAGGAATTCAGCAACGACTTCGCCCTGTGGGAGGCCGCTCAAT	1740
Db	1872	CATCACTAATTTTCTCAAGGAATTCAGCAACGACTTCGCCCTGTGGGAGGCCGCTCAAT	1931
Qy	1741	GCCCAATCAGAGATTACCAAAAGATTCGGGATGTCATGCTGAAGAACAATTCAGGGCATG	1800
Db	1932	GCCCAATCAGAGATTACCAAAAGATTCGGGATGTCATGCTGAAGAACAATTCAGGGCATG	1991
Qy	1801	AAGCACTGGCGCTCACCTGTGGAAGCACAGCAGGCGCTTGGAGGCCCTTGAGAAATGGA	1860

Db	1992	AAGCACCTGCGGGCTCACCTGTGGGAAGCACACGAGAGGCCCTTGGAGAGCCCTGGAGAAATGGGA	2055
Qy	1861	ATCAAGAGCTCCCGCGGGCTGGAGAACCTTCTGCAGAGACTTTTGAGCTGCGAGAAGGTGTGT	1920
Db	2052	ATCAAGAGCTCCCGCGGGCTGGAGAACCTTCTGCAGAGACTTTGAGCTGCGAGAAGGTGTGT	2111
Qy	1921	TACCTAGCGCTCAACACCTTCCCTGCGGGCACTGCACCGGCTCATGCACTACAACGAG	1980
Db	2112	TACCTAGCGCTCAACACCTTCCCTGCGGGCACTGCACCGGCTCATGCACTACAACGAG	2171
Qy	1981	GTCTTGAGCGGCTGTGCAAAACACACCCCGCGAGCCAGCCGACTTCAGGGACTGCCTGA	2040
Db	2172	GTCTTGAGCGGCTGTGCAAAACACACCCCGCGAGCCAGCCGACTTCAGGGACTGCCTGA	2231
Qy	2041	GCCGCTTTGGCAGAGATCACGGAGATGTGGCACAGCTCCACGGTACGATGATCAAGATG	2100
Db	2232	GCCGCTTTGGCAGAGATCACGGAGATGTGGCACAGCTCCACGGTACGATGATCAAGATG	2291
Qy	2101	GAGAAATTTCCAGAAGCTGCACGAACCTCAAGAAGAATTTGATTGGCATTGACAATCTTTGTG	2160
Db	2292	GAGAAATTTCCAGAAGCTGCACGAACCTCAAGAAGAATTTGATTGGCATTGACAATCTTTGTG	2351
Qy	2161	GTTCGGGAAGGAGTTCATCGTCTGGGCGACGCTCAGCAAGCTCTCGGGGAAGGGGCTC	2220
Db	2352	GTTCGGGAAGGAGTTCATCGTCTGGGCGACGCTCAGCAAGCTCTCGGGGAAGGGGCTC	2411
Qy	2221	CAGCAGCGCATGTTCTTCTGTTCAACGACGCTCCTGCTATACAGAGCCGGGGGCTCAGC	2280
Db	2412	CAGCAGCGCATGTTCTTCTGTTCAACGACGCTCCTGCTATACAGAGCCGGGGGCTCAGC	2471
Qy	2281	GCTTCCAATCAGTTTAAAGTCCACGGGCAAGCTCCCGGCTCTATGGCATGACGATTGAGGAG	2340
Db	2472	GCTTCCAATCAGTTTAAAGTCCACGGGCAAGCTCCCGGCTCTATGGCATGACGATTGAGGAG	2531
Qy	2341	AGCGAAGACGA 2351	
Db	2532	AGCGAAGACGA 2542	
RESULT 3			
AAS64760			
ID	AAS64760 standard; cDNA; 3187 BP.		
AC	AAS64760;		
CC			
DT	13-FEB-2002 (first entry)		
XX	DNA encoding novel human diagnostic protein #564.		
DE	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
KW	Homo sapiens.		
OS	WO200175067-A2.		
XX	11-OCT-2001.		
PN	30-MAR-2001; 2001WO-US08631.		
XX	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX	(HYSE-) HYSEQ INC.		
PA	Drmanac RT, Liu C, Tang YT;		
XX	WPI; 2001-639362/73.		
PI	P-PSDB; ABG00573.		
DR	New isolated polynucleotide and encoded polypeptides, useful in		
XX	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT			

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess

PT biodiversity -  
XX Claim 1; SEQ ID No 564; 103pp; English.  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3187 BP; 810 A; 874 C; 876 G; 627 T; 0 other;

Query Match 99.3%; Score 2335.2; DB 23; Length 3187;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2348; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 CCTCTCAGGAAACTCGTGTCCATCAAAATCCAGATCGTGGATGACACCCAGGAGGCATTT 60  
DB 151 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATCGTGGATGACACCCAGGAGGCATTT 210  
QY 61 GAAGTCCCAAAAGAGCTCCTCGGAAGGTGCTGCTGGATGCAATTTGCAACCACTCAAC 120  
DB 211 GAAGTCCCAAAAGAGCTCCTCGGAAGGTGCTGCTGGATGCAATTTGCAACCACTCAAC 270  
QY 121 CTCGTGGAAGGTGACATATTTGGCCCTCGAGTTTCCTGATGATCACAATAAGATCAGCGTGG 180  
DB 271 CTCGTGGAAGGTGACATATTTGGCCCTCGAGTTTCCTGATGATCACAATAAGATCAGCGTGG 330  
QY 181 CTGGATCTCCTTAAACCCATTTGGAACAGATTAGAAGGCCAAGCACCGTTGTTGTTAAG 240  
DB 331 CTGGATCTCCTTAAACCCATTTGGAACAGATTAGAAGGCCAAGCACCGTTGTTGTTAAG 390  
QY 241 TTTGTGGTGAAATTTCTTCCGCTGACCAACACACAACTCCAAGAAGAACTCACAAGGTAC 300  
DB 391 TTTGTGGTGAAATTTCTTCCGCTGACCAACACACAACTCCAAGAAGAACTCACAAGGTAC 450  
QY 301 CTGTTCCGCTGCGAGTGAAGCAGGACTTGGCTCAAGCAGGTTGACGTGTAATGACACC 360  
DB 451 CTGTTCCGCTGCGAGTGAAGCAGGACTTGGCTCAAGCAGGTTGACGTGTAATGACACC 510  
QY 361 AGCGCAGCTCTCTTGATTTTACACATTTGCAATCTGAGATTGGGATTTTGATGAAGCC 420  
DB 511 AGCGCAGCTCTCTTGATTTTACACATTTGCAATCTGAGATTGGGATTTTGATGAAGCC 570  
QY 421 TTGGCAGAGAGACATTTAGCAAAAAATAAATACATACCTCAGCAAGACGCACTAGAGGAC 480  
DB 571 TTGGCAGAGAGACATTTAGCAAAAAATAAATACATACCTCAGCAAGACGCACTAGAGGAC 630  
QY 481 ARAATCGTGAATTTTACCATTACCATTTGGCAACACACCAGCAAGATTCAGATTTCCAG 540  
DB 631 ARAATCGTGAATTTTACCATTACCATTTGGCAACACACCAGCAAGATTCAGATTTCCAG 690  
QY 541 CTCCTAGAGATTGCCCGCTCGGTAGAGATGATGGAATCCGGTTGCACCCGGCCCAAGGAC 600  
DB 691 CTCCTAGAGATTGCCCGCTCGGTAGAGATGATGGAATCCGGTTGCACCCGGCCCAAGGAC 750  
QY 601 AGGGAAGGCACGAAGATCAATCTGGCCGTTGCCAACACGCGGAATTCAGTGTTCAGGGT 660

DB 751 AGGGAAGGCACGAAGATCAATCTGGCCGTTGCCAACACGGGAATTCAGTGTTCAGGGT 810  
QY 661 TTCACCTAAGATCAATGCCTTCAACTGGGCCAAGGTCGGAGCTGAGCTTCAAGAGAAG 720  
DB 811 TTCACCTAAGATCAATGCCTTCAACTGGGCCAAGGTCGGAGCTGAGCTTCAAGAGAAG 870  
QY 721 CGCTTTCTCATCAAGCTCCGGCCAGATGCCAATAGTGCCTACCAAGATACCTTTGGAATTC 780  
DB 871 CGCTTTCTCATCAAGCTCCGGCCAGATGCCAATAGTGCCTACCAAGATACCTTTGGAATTC 930  
QY 781 CTGATGCGCAGTCGGGATTTCTGCAAGTCTCTTCTGAAAATCTGTGTGAACATCATGCC 840  
DB 931 CTGATGCGCAGTCGGGATTTCTGCAAGTCTCTTCTGAAAATCTGTGTGAACATCATGCC 990  
QY 841 TTCCTTTAGACTTTTTCGAAGGCCCAACCAAGCCCAAGCCCGTCTCTTTAGCCGGGG 900  
DB 991 TTCCTTTAGACTTTTTCGAAGGCCCAACCAAGCCCAAGCCCGTCTCTTTAGCCGGGG 1050  
QY 901 TCATCATTTTCGGTTCAGTGGTCCGACTCAGAAGCAGGTTCTCGACTATGTTAAAGAAGGA 960  
DB 1051 TCATCATTTTCGGTTCAGTGGTCCGACTCAGAAGCAGGTTCTCGACTATGTTAAAGAAGGA 1110  
QY 961 GGACATAAGAAGGTGCAGTTTGAAGGAAGCACAGCAAGATTCTATPCGGAGCCCT 1020  
DB 1111 GGACATAAGAAGGTGCAGTTTGAAGGAAGCACAGCAAGATTCTATPCGGAGCCCT 1170  
QY 1021 GCTTCACAGCTACAGAACTGAATTCGGAAGTCTCGAGCAGTCTCAGCAGACACACAG 1080  
DB 1171 GCTTCACAGCTACAGAACTGAATTCGGAAGTCTCGAGCAGTCTCAGCAGACACACAG 1230  
QY 1081 CTTTACATTTGGAGA-AGGTGCCGAATCTCCAGGGGCCAGAGCTGCCGGCAGGAAAGGA 1139  
DB 1231 CTTTACATTTGGAGAGAGTGCCTGCAATTTCCAGGGGCCAGAGCTGCCGGCAGGAAAGGA 1290  
QY 1140 ACCGAAGTTCCTCGCCGGGAGCCGGGTGCGACCCGAGCCCTGCGCCCGAGGAGAAGCC 1199  
DB 1291 ACCGAAGTTCCTCGCCGGGAGCCGGGTGCGACCCGAGCCCTGCGCCCGAGGAGAAGCC 1350  
QY 1200 CCGGGTAAACAGCAGCGGACGCGGCCCTCGGCCGCCACCGAGGAGAAGAGGAGGT 1259  
DB 1351 CCGGGTAAACAGCAGCGGACGCGGCCCTCGGCCGCCACCGAGGAGAAGAGGAGGT 1410  
QY 1260 CGTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCGCCGAGCCCAAGCAGAGCTCCCT 1319  
DB 1411 CGTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCTCGCAGCCCAAGCAGAGCTCCCT 1470  
QY 1320 GACTGCGAGTCTCCTTTCCGAGTCTGTGTGAATCTCGCAGGGGGAGTGGCCCTGC 1379  
DB 1471 GACTGCGAGTCTCCTTTCCGAGTCTGTGTGAATCTCGCAGGGGGAGTGGCCCTGC 1530  
QY 1380 CAACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAG 1439  
DB 1531 CAACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAG 1590  
QY 1440 CCGCTGCTGAATGACACAGGCTTGCCTCCGAGCGAGCAGATGAGGATGAGGCGCGAGGAA 1499  
DB 1591 CCGCTGCTGAATGACACAGGCTTGCCTCCGAGCGAGCAGATGAGGATGAGGCGCGAGGAA 1650  
QY 1500 GAGATTCCCAACTGATAAAGCGTACTTCTATAGTAAGGAAGTCTTACCAAGCAGGCAAC 1559  
DB 1651 GAGATTCCCAACTGATAAAGCGTACTTCTATAGTAAGGAAGTCTTACCAAGCAGGCAAC 1710  
QY 1560 ATATCTGAAGGATCTCGAAGTATATCACTTCGTGGTTTCAGAGCAGCAGTGAAGCAAGGA 1619  
DB 1711 ATATCTGAAGGATCTCGAAGTATATCACTTCGTGGTTTCAGAGCAGCAGTGAAGCAAGGA 1770  
QY 1620 CGCCATGCCGGAAGCAGTGAAGGCTCTATATTTCCCGAATTTTGAACCTTTGCACAAAT 1679  
DB 1771 CGCCATGCCGGAAGCAGTGAAGGCTCTATATTTCCCGAATTTTGAACCTTTGCACAAAT 1830  
QY 1680 TCATACTAATTTTCTCAAGGAATTCAGCAAGCAGTTCGCCCTGTGGGAAGGCCGCTCAA 1739

Db 1831 TCATACTAATTTTCTCAAGAAATTTAGCAACGACTTGGCCCTGTGGGAAGCGCGCTCAAA 1890  
QY TGCCCAAATCAGAGATTACCAAGAATCGCGATGTGTCATGCTGAAGAACATTCAGGGCAT 1799  
Db 1891 TGCCCAAATCAGAGATTACCAAGAATCGCGATGTGTCATGCTGAAGAACATTCAGGGCAT 1950  
QY 1800 GAAGCACCTTGGCGGCTCACCTGTGTGAAGCACAGCGAGGCGCTTGGAGGCGCTTGAGAAATGG 1859  
Db 1951 GAAGCACCTTGGCGGCTCACCTGTGTGAAGCACAGCGAGGCGCTTGGAGGCGCTTGAGAAATGG 2010  
QY 1860 AATCAGAGCTCCCGGCGCTGAGAACATTCGACAGAGATTTGAGCTGCAGAAAGTGTG 1919  
Db 2011 AATCAGAGCTCCCGGCGCTGAGAACATTCGACAGAGATTTGAGCTGCAGAAAGTGTG 2070  
QY 1920 TTACCTACCGCTCAACACCTTCTCTCTCGGCGACTGCACCGGCTCATGCATACAAAGCA 1979  
Db 2071 TTACCTACCGCTCAACACCTTCTCTCTCGGCGACTGCACCGGCTCATGCATACAAAGCA 2130  
QY 1980 GGTCTTGGAGCGGCTGTGCAAAACACACCGCGGAGCCACGCGGACTTCAGGGACTTGGCG 2039  
Db 2131 GGTCTTGGAGCGGCTGTGCAAAACACACCGCGGAGCCACGCGGACTTCAGGGACTTGGCG 2190  
QY 2040 AGCCGCTTTGGCAGAGATCAGGAGATGTGTGCACAGCTCCACGGTACGATGATCAAGAT 2099  
Db 2191 AGCCGCTTTGGCAGAGATCAGGAGATGTGTGCACAGCTCCACGGTACGATGATCAAGAT 2250  
QY 2100 GGAGAAATTTCCAGAACTGCAGAACTCAAGAAAGATTTGATTGGCATTGCAAACTCTTCT 2159  
Db 2251 GGAGAAATTTCCAGAACTGCAGAACTCAAGAAAGATTTGATTGGCATTGCAAACTCTTCT 2310  
QY 2160 GGTCCGGAAGGAGTTTCATCCGTTGGCGAGCCCTCAGCAAGCTCTCGGGGAAGGGGCT 2219  
Db 2311 GGTCCGGAAGGAGTTTCATCCGTTGGCGAGCCCTCAGCAAGCTCTCGGGGAAGGGGCT 2370  
QY 2220 CCAGCAGCGCATGTTCTTCTGTTCAACAGCTCTGCTATACAGAGCGGGGCGCTGAC 2279  
Db 2371 CCAGCAGCGCATGTTCTTCTGTTCAACAGCTCTGCTATACAGAGCGGGGCGCTGAC 2430  
QY 2280 GGCTCCAACTAGTTTAAAGTCCACGGGAGCTCCGCGCTATGGCATGACGATTGAGCA 2339  
Db 2431 GGCTCCAACTAGTTTAAAGTCCACGGGAGCTCCGCGCTATGGCATGACGATTGAGCA 2490  
QY 2340 GAGCGAAGACGA 2351  
Db 2491 GAGCGAAGACGA 2502  
RESULT 4  
AAC98992  
ID AAC98992 standard; cdNA; 3094 BP.  
XX  
AC AAC98992;  
XX  
DT 09-MAR-2001 (first entry)  
DE  
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.  
XX  
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW pulmonary; cardiovascular; renal; proliferative; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20005320-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05989.  
XX

PR 12-MAR-1999; 990S-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2000-579444/54.  
DR P-PSDB; AAB54227.  
XX  
PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
XX  
PS Claim 1; Page 664-665; 1379pp; English.  
XX  
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 3094 BP; 736 A; 849 C; 755 G; 749 T; 5 other;  
Query Match 32.4%; Score 762; DB 21; Length 3094;  
Best Local Similarity 100.0%; Pred. No. 3.2e-205;  
Matches 762; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1590 GTGGTTTCAGACAGTGCAGCAAGAGAGACGCGCATGCCGAGGAGCTCAAAAGTCTCAT 1649  
Db 62 GTGGTTTCAGACAGTGCAGCAAGAGAGAGCGCCATGCCGAGGAGCTCAAAAGTCTCAT 121  
QY 1650 ATTCCCGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAAATTCGCA 1709  
Db 122 ATTCCCGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAAATTCGCA 181  
QY 1710 ACGACTTGCCCTGTGGGAAGGCGCTCAATGCCCAATCAGAGATTACCAAGAAATTCGG 1769  
Db 182 ACGACTTGCCCTGTGGGAAGGCGCTCAATGCCCAATCAGAGATTACCAAGAAATTCGG 241  
QY 1770 CGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACTGTGGAAGCA 1829  
Db 242 CGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACTGTGGAAGCA 301  
QY 1830 CAGCGAGGCTTTGGAGGCGCTTGAGAATGGAATCAAGAGCTCCCGGCGGCTGGAGAACTT 1889  
Db 302 CAGCGAGGCTTTGGAGGCGCTTGAGAATGGAATCAAGAGCTCCCGGCGGCTGGAGAACTT 361  
QY 1890 CTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGG 1949  
Db 362 CTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGG 421  
QY 1950 GCCACTGACCGGCTCATGCTACTACAAGCAGGTCTCTGGAGCGGCTGTGCAACACCCACC 2009  
Db 422 GCCACTGACCGGCTCATGCTACTACAAGCAGGTCTCTGGAGCGGCTGTGCAACACCCACC 481  
QY 2010 GCCGAGCCACGCCGACTTTCAGGAGCTTGCAGCGCGCTTTTGGCAGAGATCACGGAGATG 2069

Db 482 GCGGAGCCAGCCGACTTCAGGAGTCGCGAGCCGCTTTGGCAGAGATCACGGAGTGT 541  
QY 2070 GGCACAGCTCCACGGTACGATCAAGATGAGAAATTTCCAGAAGCTGCAGAACTCAA 2129  
Db 542 GGCACAGCTCCACGGTACGATCAAGATGAGAAATTTCCAGAAGCTGCAGAACTCAA 601  
QY 2130 GAAAGATTGATTGGCATGACAATCTTGTTCCGGGAAGGGATTCATCCGCTCTGGG 2189  
Db 602 GAAAGATTGATTGGCATGACAATCTTGTTCCGGGAAGGGATTCATCCGCTCTGGG 661  
QY 2190 GAGCCTCAGCAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCCTGTTCAACGA 2249  
Db 662 GAGCCTCAGCAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCCTGTTCAACGA 721  
QY 2250 CGTCTGCTATACAGCAGCGCGGGGCTGACGCGCTCCCAATCAGTTTAAAGTCCACGGGCA 2309  
Db 722 CGTCTGCTATACAGCAGCGCGGGGCTGACGCGCTCCCAATCAGTTTAAAGTCCACGGGCA 781  
QY 2310 GTCCTCGCTCTATGCTATGACGATGAGGAGAGCGAAGACGA 2351  
Db 782 GTCCTCGCTCTATGCTATGACGATGAGGAGAGCGAAGACGA 823  
RESULT 5  
AAS64761/c  
ID AAS64761 standard; cDNA; 1824 BP.  
XX AC AAS64761;  
XX  
XX 13-FEB-2002 (first entry)  
DT  
DE DNA encoding novel human diagnostic protein #565.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX 31-MAR-2000; 2000US-0540217.  
PR  
XX 23-AUG-2000; 2000US-0649167.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX WPI; 2001-639362/73.  
DR  
XX P-PSDB; ABG00574.  
DR  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 565; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1824 BP; 409 A; 477 C; 486 G; 452 T; 0 other;  
Query Match 10.8%; Score 254; DB 23; Length 1824;  
Best Local Similarity 100.0%; Pred. No. 3.5e-61;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1060 CAGTCTCAGCAGAGCAGCAGCCTTACATTTGGAGAAGGTTCCTCCAGGGGGCCAG 1119  
Db 1623 CAGTCTCAGCAGAGCAGCAGCCTTACATTTGGAGAAGGTTCCTCCAGGGGGCCAG 1564  
QY 1120 AGCTGCCGCGAGGAAAGAACCCGAGGTTTCCGCGGGGAGCGGGGTTCGCACCCGAGC 1179  
Db 1563 AGCTGCCGCGAGGAAAGAACCCGAGGTTTCCGCGGGGAGCGGGGTTCGCACCCGAGC 1504  
QY 1180 CCTGCGCGGAGGAGAGCCCGCGGGTTAACAGCAGCGGAGCGCCCTTCGGGCGCC 1239  
Db 1503 CCTGCGCGGAGGAGAGCCCGCGGGTTAACAGCAGCGGAGCGCCCTTCGGGCGCC 1444  
QY 1240 ACGGAGGAAGAGGAGGAGTCTTTAAGGATAGGAGCCAGCAGAGTAAACCTCAGCCCCCG 1299  
Db 1443 ACGGAGGAAGAGGAGGAGTCTTTAAGGATAGGAGCCAGCAGAGTAAACCTCAGCCCCCG 1384  
QY 1300 CAGCCAAGCACAGG 1313  
Db 1383 CAGCCAAGCACAGG 1370  
RESULT 6  
ABL02335  
ID ABL02335 standard; cDNA; 1220 BP.  
XX  
XX AC ABL02335;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 1487.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX 11-JUL-2000; 2000US-0614150.  
PR  
XX (PEKE ) PE CORP NY.  
PA  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR  
XX P-PSDB; ABB58232.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions  
XX



PS Claim 1; SEQ ID NO 1487; 2lpp + Sequence Listing; English.

xx The invention relates to an isolated nucleic acid detection reagent

cc capable of detecting 1000 or more genes from Drosophila. The invention is

cc useful in developmental biology and in elucidating cell signalling and

cc cell-cell interactions in higher eukaryotes for the development of

cc insecticides, therapeutics and pharmaceutical drugs. The invention

cc discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

cc sequences (ABL01840-ABL16175) and the encoded proteins

cc (AB57737-AB572072).

cc The sequence data for this patent did not form part of the printed

cc specification, but was obtained in electronic format directly from WIPO

cc at ftp.wipo.int/pub/published\_pct\_sequences.

xx

sq Sequence 1220 BP; 309 A; 327 C; 329 G; 255 T; 0 other;

Query Match 9.78; Score 228; DB 23; Length 1220;

Best Local Similarity 55.18; Pred. No. 6.6e-54;

Matches 557; Conservative 0; Mismatches 435; Indels 18; Gaps 5;

Qy 7 GGAATACTCGTTCATCAAAATCCAGATGCTGGATGACACCCAGGAGGCAATTTGAAGTT 66

Db 198 GGCRAAAGCTGATGTTTCGATCCAGATGCTGGAGACTGATCACCATTGTTCCAGTA 257

Qy 67 CCACAAAGAGTCTCTGGGAAGTGTCTGGATGACGTTTCCAAACCACTCAACCTCGTG 126

Db 258 CAGGCTAAAGCACTGGGCGCGTGTGTTGAACAGGATGCTGCGTCAATTTACTG 317

Qy 127 GAAGTGTACTATTTGGCTCGAGTTTCTTGATCA---CAAAAGATCAGCGTGGCTG 183

Db 318 GAGGCGCACTACTTGGCTTGGATGACCAAGTCTCCACATACATACTGGCTG 377

Qy 184 GATCTCTTAAACCACTTGTGAACAGATTAAGAGGCGCAAGCGTGTGTTAAAG--- 240

Db 378 GATCTGGAGAGCAATGAACCGCAGTGGGCTCTCTTATCGATCCAGTCTGGC 437

Qy 241 TTTGTGGTGAATCTTTCCGCTGACACACAACTCCAAAGAACTCACAAGGTAC 300

Db 438 TTCTGCATCAAGTTCTATACACCGGATCCAGCGCAATTTGGAAGAGGATACACAGGTAT 497

Qy 301 CTGTTGCGCTGCAAGTGAAGCAAGTGTGCTCAAGCGAGTTCAGTGTGAATGACAC 360

Db 498 TTGTTCTGCTGCAATCAAGAGAGCTGGCCAGCGAGCTGCGAGTGAACGATAC 557

Qy 361 AGCGCAGCTCTTGAATTCACATTTGCAATCTGAGATTTGGGATTTTGAAT---GAAG 418

Db 558 ACGGCGCTTAATGGCAAGTACATTTACAGCGTATCGGCGAGTCTGTTCCGGAG 617

Qy 419 CTTGGACAGAGACTTTAGCAAAATAAATACATACCT-----CAGCAAGACGCA 471

Db 618 GATTATCCCGACCACTATCTGTCTCTGATGCTTTGTGCCCAACACGAGCGCCACC 677

Qy 472 CTAGAGGACAAATCGTGAATTTACCATACACATGACATGACCAACACAGCAGATCA 531

Db 678 ATGACGGAAGATTTATGAGAACCAACAAAGGATGTTGTTCAATCCCGCGGAGCG 737

Qy 532 GATTTCAGCTCTAGAGATGCGCGTTCGCTGAGATGATGGAATCCGTTGGCACCG 591

Db 738 GACCTTAACCTTTGGACACGCTCGGAGTGTGAGCTGTATGCGTGAATGATCCG 797

Qy 592 GCCAAGGACAGGAGGACCAAGATCAATCTGGCGGTGGCCACACAGGGAATTTAGTG 651

Db 798 GCGAAGGATGTGGAAGGGTGGCGTTAATCTGGCTGTGGGCCACATGGGCAATCACAGTC 857

Qy 652 TTTCAGGTTTCACTAAGATCAATGCTTCAACTGGGCGCAAGGTGCGGAAGCTGAGCTTC 711

Db 858 TTCCAGAACATACCGGATCAACACCTTCTGTTGGGTAGATGACGAATTTCTTTC 917

Qy 712 AAGAGGACGCGTTTCTCATCAAGTCCGCGCAGATGATGCGTACAGGATACAC 771

Db 918 AAGCGCAAGGATTCCTGTTCAACTGATCGGAGGATGATGATATACAGGATAC 977

Qy 772 TTGGAATTCGTAGTGGCCAGTGGGATTTCTGCAAGTCTTCTGGAAATCTGTGTGAA 831

Db 978 GTGAGTCTCTTCTCGAGGGTCGCAAGAGTGCAGAACTTCTGCAAAAATGCGTGCA 1037

Qy 832 CATCATCGCTCTTTAGACTTTTGAAGAGCCAA---ACCAAGCCCAAGCCGCTGCTC 888

Db 1038 AATCAGGATTTCTTCGATGACATGCGGTACAAATACGCCAGCGCAAACTCGCGTT 1097

Qy 889 TTTAGCGGGGCTCATCATTTTCGTTTCAGTGTGCGGACTCAGAAAGAGTTCCTGACTAT 948

Db 1098 CTCTCGCGGGTAGTTTCATTCCTGCTATAGCGGAAACCCAGAGGATATCGAGTTC 1157

Qy 949 GTTAAAGAGGAGGACATCAAGAGGTGCGAGTTTGAAGGAAGCAGCAGCA 998

Db 1158 GTTCGCAAAATTTATGTGAAGCGCCAAACTTCCAAAGGTATATAAACGAA 1207

RESULT 7

AAA42150/c

ID AAA42150 standard; cDNA; 241 BP.

XX

AC AAA42150;

DT 21-AUG-2000 (first entry)

XX

Human secreted expressed sequence tag SEQ ID NO:890.

Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;

expressed sequence tag; EST; probe; chemotactic; proliferative;

immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

thrombolytic; antiinflammatory; cytosolic; antibacterial; antifungal;

antiviral; osteopathic; antiasthmatic; vulnerary; antiparkinsonian;

antitumor; neuroprotective; neurotropic; antipsoriatic;

cerebroprotective; anticonvulsant; antidepressant; gene therapy;

vaccine; autoimmune disorder; multiple sclerosis; allergic condition;

insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;

lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;

central nervous system disorder; Alzheimer's disease; stroke;

Parkinson's disease; Huntington's disease; coagulation disorder;

haemophilia; thrombosis; inflammatory disorder; Crohn's disease;

tumour; infection; depression; psoriasis; ss.

XX

Homo sapiens.

OS

XX

WO200021990-A1.

PN

XX

20-APR-2000.

XX

15-OCT-1999; 99WO-US24205.

PF

XX

15-OCT-1998; 98US-0104435.

PR

XX

(GEMY ) GENETICS INST INC.

PA

XX

Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M;

PI

XX

WPI; 2000-317937/27.

DR

XX

Isolated polynucleotides, and encoded proteins, comprising secreted

expressed sequence tags (sESTs), useful for treating various disorders

such as autoimmune, infectious, and central nervous system disorders -

XX

Claim 1; Page 354; 618pp; English.

PS

XX

AAAA1261 to AAA43419 represent specifically claimed secreted expressed

sequence tags (sESTs), isolated from human, mouse, xenopus and rat

tissue sources. The sESTs can have a range of activities depending on

the tissues they were isolated from. The activities include:

CC chemotactic; proliferative; immunomodulatory; haematopoietic;

CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;

CC cytosolic; antibacterial; antifungal; antiviral; antidiabetic;

CC antiasthmatic; vulnerary; antitumor; osteopathic; neuroprotective;

CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;





Db 1464 AATTATGATTGGAGTGATGTCAGGAGGAATTCGTATTATATAAGAACAGGGTACGAATGAA 1523  
 QY 675 TGCCTTTAACTCGCGCAAGGTCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTTCATCAA 734  
 Db 1524 TACCTTTCCATGGTTGAAGATTGTAATAATTTCTTTTAAGTGCACACAGATTTTATTCA 1583  
 QY 735 GCTCGGGCAGATGCCAATAGTCGTACCAAGATACCTTGGGAATTCCTGATGGCCGTCG 794  
 Db 1584 ACTTAAAGAAATATGCATGAATCAGAGAACATTAATTTGGGATTTAATATGTCGAATTA 1643  
 QY 795 GGATTTCTGCAAGTCTTCTGGAATCTGTTGAAACATCATGCTTCTTTAGACTTTT 854  
 Db 1644 CAGACATGTAAATAATTTGTGGAAAGCATGTGTAGAATCATACATCTTCCGTTTGA 1703  
 QY 855 TGAAGAGGCCAACCAAGCCCCAA-----GCCCTCTCTTTAGCGGGGTCTCATATT 908  
 Db 1704 CAGACCATCTCACCTCAACAGAAATTTTTCGACATATTTTACATTAAGTTCAAAAT 1763  
 QY 909 TCGGTTCAAGTGTGGAGTCTGAGACAGGTTCTCGACTATGTTAAAGAGGAGGACATAA 968  
 Db 1764 CCGGTACTGTGGAGAACTGAAGTCCAATCAGTTCAGTATGSCAAAGAAAGGCAATAA 1823  
 QY 969 GAAGTGCAGTTTGAAGGAACACAGCAAG 999  
 Db 1824 AGACAGGGTATTTGCAAGATCCCAAGTAAG 1854

## RESULT 10

ABK70019  
 ID ABK70019 standard; cDNA; 6263 BP.

AC ABK70019;

XX 15-JUL-2002 (first entry)

DE Human cDNA for the KIA00338 gene.

XX Tax interacting protein clone 2; TIP-2; GIPC; antigen; tumour;  
 KW breast cancer; prostate cancer; monoclonal antibody; 27.B1; 27.F7;  
 KW thyroid cancer; tetanus toxin; infection; HIV; hanta virus; Ebola;  
 KW human immunodeficiency virus infection; enzyme dysfunction;  
 KW hormone dysfunction; autoimmune disease; lupus; rheumatoid arthritis;  
 KW immune dysfunction; septicaemia; human; ss; gene.

XX Homo sapiens.

OS WO200222851-A2.

PN 21-MAR-2002.

XX 18-SEP-2001; 2001WO-US29242.

PF 18-SEP-2000; 2000US-0664958.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

XX Trakht I, Canfield R, Kalantarov G, Rudchenko S;

PI WPI: 2002-362353/39.

DR P-PSDB; ABG34123.

XX New monoclonal antibody which specifically binds and forms complex with  
 PT TIP-2 antigen located on surface of human cancer cells, useful for  
 PT diagnosing and treating cancer in a human subject -

XX Example 1; Fig 32A-F; 276pp; English.

XX The invention relates to a monoclonal (mAb) antibody which specifically  
 CC binds and forms a complex with TIP-2 (Tax interacting protein clone 2,  
 CC also known as GIPC) antigen located on the surface of human cancer cells,  
 CC where mAb binds to the same antigen as monoclonal antibody 27.B1  
 CC or 27.F7 produced by hybridoma 27.B1 or 27.F7 of ATCC Designation No.  
 CC PTA-1599 or 1598, respectively. The mAb is useful for detecting TIP-2

CC antigen bearing cancer cells, for diagnosing cancer in a subject by  
 CC detecting TIP-2 antigen-bearing cancer cells, for in vivo diagnosis of  
 CC cancer in a subject, for delivering exogenous material to TIP-2 antigen-  
 CC bearing cancer cells of a human subject, for treating cancer in a human  
 CC subject, for inducing apoptosis of TIP-2 antigen bearing cells, for  
 CC immunohistochemical screening of a tissue section from a tumour sample  
 CC for the presence of TIP-2 antigen bearing cancer cells, for detecting the  
 CC presence of TIP-2 antigen in biological fluid, and for monitoring  
 CC progression of cancer, where the cancer cells are TIP-2 antigen-bearing  
 CC cancer cells, in a subject. A composition comprising the mAb is useful  
 CC for treating or preventing a condition in a subject who previously  
 CC exhibited the condition, where the condition is associated with cancer  
 CC (thyroid, breast or prostate cancer), tumour (benign), toxin (tetanus,  
 CC anthrax, botulinum, snake venom or spider venom), infectious agent (such  
 CC as Hanta virus, HTLV (not defined) I, HTLV II, Human immunodeficiency  
 CC virus, HIV herpes virus, influenza, Ebola, human papilloma virus,  
 CC Staphylococcus, Streptococcus, Klebsiella, Escherichia coli, anthrax or  
 CC Cryptococcus), enzyme dysfunction (hyperactivity or overproduction of the  
 CC enzyme), hormone dysfunction (hyperactivity or overproduction of the  
 CC hormone), autoimmune disease (lupus, thyroiditis, graft versus host  
 CC disease, transplantation rejection or rheumatoid arthritis), immune  
 CC dysfunction (CD3 or CD4 mediated), viral antigen, bacterial antigen,  
 CC eukaryotic antigen, rejection of a transplanted tissue, or the condition  
 CC is septicaemia, sepsis, septic shock, viraemia, bacteraemia, fungaemia.  
 CC The present sequence encodes an additional human cancer associated marker  
 CC which can be used in the production of monoclonal antibodies.

XX Sequence 6263 BP; 1516 A; 1763 C; 1591 G; 1391 T; 2 other;

Query Match 7.7%; Score 180.4; DB 24; Length 6263;

Best Local Similarity 52.9%; Pred. No. 5.4e-40;

Matches 485; Conservative 0; Mismatches 416; Indels 15; Gaps 4;

QY 35 TCGTGGATGACACCCAGGAGGCAATTTGAAGTTCACAAAGAGCTCTGGGAGGTGCTGC 94

Db 469 TGCCTTATGCGCTCGGAGTATGATGTGAGGTGGAGAAACATGGCCGGCCAGGTGCTGT 528

QY 95 TGGATGCGAGTTTGGCAACACCTCAACCTCGTGGAGGTGACTATTTTGGCCTCGAGTTTC 154

Db 529 TTGACCTGGTGTGTGAACACCTCAACCTCTAGAGAGGACTACTTCGGCCTGACCTTCT 588

QY 155 CTGATCACAAAAGATCACGGTGTGGTGATCTCCTAAAACCCATTGTGAAACAGATTA 214

Db 589 GTGATGCTGACAGCCAGAGAAGAACTGGCTGGACCCCTCCAAAGGAGATCAAGAGCAGATCC 648

QY 215 GAAGGCCAAAGCACGTTGTTGTTAGTTTGTGGTGAATTTCTTCGGCTGACCCACACAC 274

Db 649 GGAG---TAGCCCTTGGAAATTTTGCTTACAGTCAAGTTCTACCCGCTGATCTCGCC 705

QY 275 AACTCCAAGAAGAACTCAACAGGTACTTTCGCGCTGCAGGTGAAGCAGGACTTGGCTC 334

Db 706 AGCTGACAGAGACATCACAGATACTACTCTGCTGCGCTGCGGGCAGACATCATCA 765

QY 335 AAGCAGGTTGACGTGTAATGACACGCGCAGCTCTCTTGATTTTACACATTTGCAAT 394

Db 766 CGGCGCGGCTGCCATGCTCCTTTGTGACGATGCCCTACTTGGGCTCCTACGCTGTGCGAG 825

QY 395 CTGAGATTGGGGATTTTGTATGAGGCTTTGGACAGAGAGCAGCTTAGCAAAAATAATACA 454

Db 826 CTGAGCTGGGTGACTATGATGCTGAGAGCATGTGGGCAACTATGTCAGCGAGCTCCGCT 885

QY 455 TACC-----TCAGCAAGACGCGCTAGAGGACAAAATCGTGAATTTTACACATAACCACA 508

Db 886 TCGCCCTTAAACAGACCCCGGAGCTGGAGAGAGATCATGGAGCTGCTATAGACATATA 945

QY 509 TTGGACAAACACACAGCAATCAGATTTTCCAGCTCTTAGAGATTTGCCCTCGGCTAGAGA 568

Db 946 GGGGATGACCCCGGAGAGAGCAGAAATCCACTTCTTAGAGAATGCCAAGAAGCTTTCCA 1005

QY 569 TGTATGGAATCCGGTTGCAACCGCGGCAAGGACAGGAGGACGAGATCAATCTGGCGG 628

Db 1005 TGTAGGAGTAGACCTGCACCATGCCAGGACTCTGAGGGCATCGACATCATGTTAGCGG 1065

QY 629 TTGCCAACACGGGAATCTAGTGTTCAGGGTTTTCACCTAAGATCAATGCCTTCAACTGGG 688  
D 1066 TTTGTGCCAATGGCGTCTCATCTACCGGACCGGCTGAGAATCAACCGCTTGGCTGGC 1125  
QY 689 CCAAGGTGGGAAGCTGAGCTTCAAGAGAGCGCTTTCTCATCAAGCTCCGGCGAGATG 748  
D 1126 CCAAGATCCTCAAGATCTCCTACAAAGAGAGTAACCTTCTATATCAAGATCGGGCTGGG 1185  
QY 749 CCAATAGTGGCTACAGGATACCTTTGGAATTCCTCATGGCCAGTCGGGATTTCTGCAAGT 808  
D 1186 AGTATGAGCAATTTGAGAGCAATTTGGCTTTAAGCTCCCAACACCGGTCAGCCAAGA 1245  
QY 809 CTTCTGGAAATCTGTGTGAACATCATGCTCTTTTAGACTTTTGAAGAGGCCCAAC 868  
D 1246 GACTGTGGAAGTCTGCATCGAGCAATACATTCTTCCGGC---TGTGTCCCTGAGC 1302  
QY 869 CAAAGCCCAAGCCGCTCTCTTTAGCCGGGGTCAATATTTCCGGTTCAAGTGGTGACTC 928  
D 1303 CCCCACCCCAAGGGCTTCTGGTGA---TGGGCTCCAAGTTCCGGTACAGTGGGAGGACCC 1359  
QY 929 AGAAGCAGGTTCTCGA 944  
D 1360 AGGCACAGACTCGCCA 1375

RESULT 11

ABL12469  
ID ABL12469 standard; cDNA; 2916 BP.  
XX  
AC ABL12469;  
DT  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31889.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO2001/1042-A2.  
XX  
PD 27-SEP-2001.  
PD  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB68366.

XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX  
PS Claim 1; SEQ ID NO 31889; 2lpp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 2916 BP; 833 A; 789 C; 718 G; 576 T; 0 other;  
  
Query Match 7.38; Score 171.4; DB 23; Length 2916;  
Best Local Similarity 51.0%; Pred. No. 1.3e-37;  
Matches 462; Conservative 0; Mismatches 431; Indels 12; Gaps 2;  
  
QY 35 TGCCTGGATGACACCCAGGAGGCATTTGAAGTTCCACAAAGAGCTCCTGGGAAGTGTCTGC 94  
D 486 TGCCTGGATGAGTCCCTTCTGACGCTGTCATGATCGCAAGCCATTTGGCCGTGATGTA 545  
QY 95 TGGATGAGTGTGCAACACACCTCAACCTCGTGAAGTGAATTTTGGCCCTCGAGTTTC 154  
D 546 TCAACTCAATTCGCGCGTCTGAACCTCATCGAAAGAGTACTTTTGGTGTGACCTATG 605  
QY 155 CTGATCACAAAAGATCACGGTGTGGCTGGATCTCTTAAACCCCATTTGTGAACAGATT 214  
D 606 AGACGCCACAGATCCGCGCAGCTTGGCTGGATCTGGAAAGCCGATATCCAAAGTTCTTCC 665  
QY 215 GAAGGCCAAAGCACGTTTGTGTTAAGTTTGTGGTGAATTTCTCCGCTGACACACAC 274  
D 666 GCACGGACACTT---GGCCTCTACCTTCGCCGTCAAGTTCTATCCCGCGGAGCATCGC 722  
QY 275 AACTCCAAAGAAAGTACCTTCAAGTGTTCGCGCTGAGTGAAGCAGGACTTGGCTC 334  
D 723 AGCTGAAGGAGGACATCACGGCTACCATTTGTGCTGCAGGTGCGCAATCACATCTCTG 782  
QY 335 AAGCAGGTTGACGTGTAATGACACCGCAGCTCTCTTGAATTTACACATTTGTGCAAT 394  
D 783 AGGCTGACTGCCCTGCACATTCCTCACCCACGCTCTGCTCGGATCTTACTTGTGTGAGT 842  
QY 395 CTGAGATTGGGGATTTTGTATGAAG-----CCTTGGACAGAGAGCATTAGCAAAA 445  
D 843 CGGAGATGGGCGACTACGATCGAGGAAATGCCACCAGGCGCTACCTGAAGGACTTCA 902  
QY 446 ATAATATACATACCTCAGCAAGACGCACTAGAGGACAAATCGTGAATTCATCACCATAACC 505  
D 903 AGATCGCTCCAAATCAGACGGCTGAGTTGGAGGATAAGGTCATGATCTTCAACAAGACCC 962  
QY 506 ACATTGCAACAAACAGCAGCAATCAGATTTCCAGCTCTCAGATATGCCCGTGGCTAG 565  
D 963 ACAAGGGACAAATCGCCCGCGAGGCTGAGCTACACTACCTGGAGAAATGCCAAGAGCTGG 1022  
QY 566 AGATGTATGGAATCCGGTTGCACCCGCGCAAGGACAGGACGAGGATCAATCTCTGG 625  
D 1023 CCAATGACGGCGTGGACTTGCATCCCGCTAAGGATTTGAGGGCGGTGACATCATGCTGG 1082  
QY 626 CCGTTGCCAACACAGGGAATTTCTAGTGTTCAGGGTTTCACTAAGATCAATGCCTTCAACT 685  
D 1083 CGTTTGTGCTCCGGTTTGTCTCTACCGCGATAAATTCGGCATCAACCGCTTTCCT 1142  
QY 686 GGGCCAGGTGCGGAAGCTGAGCTTCAAGAGAGAGCGGTTTCTCATCAAGCTCCGGCCAG 745  
D 1143 GGGCCAGGATTTCTGAAGATCTCTACAGCGCCACCATTTCTATACATCAAGATCCGACCGG 1202  
QY 746 ATGCCAATAGTCCGTACAGGATACCTTGAATTCCTCATGATGCCAGTCGGGATTTCTGCA 805  
D 1203 GTGAATTCGACAGTACGAATCCACCATTTGGCTTTAAGCTGGCCCAACCATCGTGGCCGCA 1262  
QY 806 AGTCCCTTCTGGAATAATCTGTGTTGAACATCATGCTCTTTAGACTTTTGAAGAGCCCA 865  
D 1263 AGAAATGTGGAATCTCTGCGTGGAGCACACACTTCTTCCGCTGATGATCCCCCGAAC 1322  
QY 866 AACCAAGCCCAAGCCGCTCTCTTTAGCCGGGGTCAATATTTCCGGTTCAAGTGTGCGA 925  
D 1323 CGGTGACAGTCCAAAGATGTTCCAGTCTTCCGGGTCCACCTATCGCTACAAGGGTTCGA 1382  
QY 926 CTCAG 930  
D 1383 CTCAG 1387

ABL18211  
ID ABL18211 standard; DNA; 3470 BP.  
AC ABL18211;  
XX  
XX 26-MAR-2002 (first entry)  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 5806.  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 5806.  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 5806; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (AB01840-ABL16175) and the encoded proteins  
CC (AB057737-AB072072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 3470 BP; 974 A; 958 C; 883 G; 655 T; 0 other;  
SQ  
Query Match 7.3%; Score 171.4; DB 23; Length 3470;  
Best Local Similarity 51.0%; Pred. No. 1.4e-37;  
Matches 462; Conservative 0; Mismatches 431; Indels 12; Gaps 2;  
QY 35 TCCTGATGACACCCAGGAGGATTTGAAGTCCACAAAGAGCTCTGGAGGTCGTGC 94  
DB 482 TGTGATGCTGCTCCCTCTGAGAGTCTCCATTCATCGCAAGGCAATGGCCGCGATGTA 541  
QY 95 TGGATCAGCTTTGCAACCACTCAACCTCGTGAAGGTGACTATTGTCCTCGAGTTTC 154  
DB 542 TCAACTCAATCTGCGCGGTCTGGAACCTCATCGAAAGGACTACTTTGGTTCACCTATG 601  
QY 155 CTGATCACAAAAGATACACGGTGTGCTGATCTCTTAAACCCATTGTGAACAGATTA 214  
DB 602 AGACGCCACAGATCCGCGACCTTGGCTGGATCTGGAGAAGCGGTATCCAAGTTCTTC 661  
QY 215 GAGGCCAAAGACAGCTGTGTTAACTGTTGGTGAATTTCTTCGCGCTGACCAACAC 274  
DB 662 GCACGACACTT---GSCCTCTCACCTTCGCGCTCAAGTTCTATCGCGGAGCCATCGC 718  
QY 275 AACTCCAGAAGAACTCAAGGTAACCTGTTCGCGTGCAGGTGAAGAGGACTTGGCTC 334  
DB 719 AGCTGAAGGAGGACATCACGCGCTACCACTTTGTGCTGCGAGGTGCGCAATGACATCCTGG 778

QY 335 AAGCGAGGTGACGTGTAATGACACACGCGCAGCTCTCTGTGATTTTACACATTGTCAAT 394  
DB 779 AGGTGCGACTGCGCTGCATTCGTCAACCCACGCTCTGCTCGGATCTACTTGTGTCAGT 838  
QY 395 CTGAGATTGGGGATTTTGTATGAAG-----CCTTGGACAGAGCAGCTTAGCAAAA 445  
DB 839 CGGAGATGGCGGACTAGCATGCGAGAGAAATGCCACACGAGGCTACCTGAAGGACTTCA 898  
QY 446 ATAAATACATACCTCAGCAAGACGCACTAGAGGACAAAATCGTGAATTTTACACATAACC 505  
DB 899 AGATCCCTCCAAATCAGACGGCTGAGTTGGAGGATGAAGTCATGGATCTTCAACAACCC 958  
QY 506 ACATTGGCAAAACACACGAGAGATCAGATTTCCAGTCTCCTAGAGATTGCCCGTGGCTAG 565  
DB 959 ACAAGGCAAAATCGCCCGGAGGCTGAGCTACACTACCTGGAGAAATGCCAAGAAGCTGG 1018  
QY 566 AGATGTATGAATCCGGTTGCACCCGCGCAAGGACAGGAGGAGGACGAGATCAATCTGG 625  
DB 1019 CCATGTACGGCGTGGACTTGCATCCCGCTAAGGATTTCTGAGGCGGTGGACATCATCTGG 1078  
QY 626 CCGTTGCCAACACGCGGAATTTCTAGTGTTCAGGGTTTCACTAAGATCAATGCCCTTCAACT 685  
DB 1079 GCGTTTGTGCTCCGGTTTGTCTCTACCGCGATAAATTCGCGCATCAACCGCTTTGCCT 1138  
QY 686 GGCCAAAGGTGCGGAAGCTGAGCTTCAAGAGGAGCGCTTTCTCATCAAGCTCCGCGCAG 745  
DB 1139 GGCCAAAGATTTCTGAAGATCTCCTCAAGCGCCACCATTTCTTACATCAAGATCCGACCG 1198  
QY 746 ATGCCAATAGTGGTACCAGGATACCTTGGAAATTTCTGATGGCCAGTCGGGATTTTCGA 805  
DB 1199 GTGAATTCGAGCATGAGAAATCCACCATTTGGCTTTAAGTGGCCAAACCATCGTGCGCCA 1258  
QY 806 AGTCCTTCTGGAAAATCTGTGTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCA 865  
DB 1259 AGAACTGTGGAATCTCTGCTGGAGCACCACACACTTCTTCCGCTGATGACCCCGAAC 1318  
QY 866 AACCAAGCCCAAGCCGCTCTTTAGCGGGGGTGCATCATTTTCGGTTCAGTGGTCGA 925  
DB 1319 CGGTACGAAAGTCCAAGATGTTCCAGTCTTCCAGTCTTCCGGTCCACCTATCGTACAAGGTCGCA 1378  
QY 926 CTCAG 930  
DB 1379 CTCAG 1383  
RESULT 13  
ABL18211  
ID ABL18211 standard; DNA; 5897 BP.  
XX  
XX ABL18211;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6106.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX









```
Db 9156 GGAGTGAGTGGCTGTGTGTTTTGGGAGGTAGCCCTGGAATTTTGCTTCACAG 9215
QY 248 TGAATTCCTTCGCGCTGACACACACACTCCAAGAAGACTCACAAGTACTGTTGG 307
Db 9216 TCAAGTTCTACCGCCTGATCCTGCCCGAGCTGACAGAAGACATCACAAGATACACTGT 9275
QY 308 CGCTGCAGGTGAAGCAGGACTTGGCTCAAGGCAGGTTGACGTGTAATGACACCACGCCAG 367
Db 9276 GCTGCACTGCGGGCAGACATCATCAGGGCCGGCTGCCATGCTCTTTGTCACGCATG 9335
QY 368 CTCTCTGATTTTCAACAATTTGCAATCTGAGATTTGGGATTTTGATGAAGCCTTGGACA 427
Db 9336 CCCTACTGGGCTCCTCAGCTGTGCAGGCTGAGCTGGGTGACTATGATGCTGAGGAGCATG 9395
QY 428 GAGAGACITTAGCAAAAATAAATACATACC-----TCAGCAAGACGACTAGAGGACA 481
Db 9396 TGGGCAACTATGTACGGAGCTCCGGTTCCGCCCTAACCCAGACCCGGGAGCTGGAGGAGA 9455
QY 482 AAATCGTGAATTTTCAACATAACACATTTGGACAACACACGACAGATCAGATTTCCAGC 541
Db 9456 GGATCATGGAGCTGCATAGACATATAGGGGATGACCCCGGAGAGCAGAAATCCACT 9515
QY 542 TCCTAGAGATTGCCCGCTCGGCTAGAGATGATGGAATCCGGTTGCACCCGGCCCAAGGACA 601
Db 9516 TCTTAGAGATGCCAAGAAGCTTTCATGTACGGAGTAGACCTGCACCATGCCAAGGACT 9575
QY 602 GGGAGGCAGAGAGATCAATCTGGCCGTTGCCAACACGGGAATTCAGTGTTCAGGGTT 661
Db 9576 CTGAGGCAATCGACATCATGTTAGGGGTTTGTGCCAATGGCCCTGCTCATCTACCCGGACC 9635
QY 662 TCACTAAGATCAATGCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTCAAGAGGAAGC 721
Db 9636 GGCTGAGAATCAACCGTTTGCCCTGSCCCNAGATCCTCAGATCTCTACAGAGAGGTA 9695
QY 722 GCTTCTCATCAAGCTCGGCCAGATGCCAATAGTCGCTACCAAGGATACCTTGAATTC 781
Db 9696 ACTTCTATATCAAGATCCGSCCTGGGAGTATCAGCAATTTGAGAGCACAAATGGCTTTA 9755
QY 782 TGATGCCAGTCGGGATTTCTGCAAGTCTCTTGGAAAATCTGTGTGAACATCATGCGCT 841
Db 9756 AGCTCCCAAAACCCAGCTAGCCCAAGAGACTGTGGAAGTCTGCATCGAGCATACAT 9815
QY 842 TCTTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCCGTCTCTTTAGCCGGGGT 901
Db 9816 TCTTCGGC---TGGTGTCCCTGAGCCCCACCCAGGGCTTCCCTGGTGA---TGGGCT 9869
QY 902 CATCATTCGGTTTCACTGTCGGACTCAGAAGCAGGTTCTCGA 944
Db 9870 CCAAGTTCGGGTACAGTGGGAGGCCCCAGGCACAGACTCGCCA 9912
```

Search completed: December 3, 2002, 18:24:07  
Job time : 476.636 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:33:49 ; Search time 71.4684 Seconds  
(without alignments)  
10088.332 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_151\_2501  
Perfect score: 2351  
Sequence: 1 ccttcaggaaactcgtgc.....attgaggagcggaagacga 2351

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56.8	2.4	4080	US-08-446-345-35	Sequence 35, Appl
2	53.8	2.3	7218	US-08-232-463-14	Sequence 14, Appl
3	42.2	1.8	4411529	US-09-103-840A-1	Sequence 1, Appl
4	40.8	1.7	4776	US-08-852-401-1	Sequence 1, Appl
5	39.6	1.7	4722	US-08-979-608A-14	Sequence 14, Appl
6	38.4	1.6	4267	US-08-949-155-51	Sequence 51, Appl
7	38.4	1.6	4267	US-09-819-964-51	Sequence 51, Appl
8	38.2	1.6	1053	5352575-6	Patent No. 5352575
9	37.6	1.6	2619	US-08-467-822-19	Sequence 19, Appl
10	37.6	1.6	2619	US-08-432-697-19	Sequence 19, Appl
11	37.6	1.6	2619	US-08-466-248-19	Sequence 19, Appl
12	37.2	1.6	562	US-09-404-879A-147	Sequence 147, App
13	37.2	1.6	44377	US-08-804-227C-7	Sequence 7, Appl
14	37.2	1.6	44377	US-08-804-198-1	Sequence 1, Appl
15	36.4	1.5	2824	US-09-010-928B-3	Sequence 3, Appl
16	36.2	1.5	822	US-09-228-986-38	Sequence 38, Appl
17	36.2	1.5	4403765	US-09-103-840A-2	Sequence 2, Appl
18	36	1.5	2214	US-08-864-038A-1	Sequence 1, Appl
19	36	1.5	3331	US-08-864-038A-2	Sequence 2, Appl
20	36	1.5	3331	US-08-864-038A-4	Sequence 4, Appl
21	36	1.5	56516	US-08-996-306-1	Sequence 1, Appl
22	36	1.5	56516	US-09-338-907-1	Sequence 1, Appl
23	36	1.5	56516	US-09-218-207-1	Sequence 1, Appl
24	36	1.5	56520	US-09-338-907-179	Sequence 179, App
25	36	1.5	56520	US-09-218-207-179	Sequence 179, App
26	35.4	1.5	2132	US-09-552-322-1	Sequence 1, Appl
27	35.4	1.5	4137	US-09-499-964-2	Sequence 2, Appl

28	35.4	1.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl
29	35.4	1.5	4411529	4	US-09-103-840A-1	Sequence 1, Appl
30	35	1.5	289	4	US-09-007-005-17	Sequence 17, Appl
31	35	1.5	289	4	US-09-244-796-17	Sequence 17, Appl
32	34.8	1.5	4822	4	US-09-484-970B-106	Sequence 106, App
33	34.6	1.5	657	4	US-09-527-345-3	Sequence 3, Appl
34	34.6	1.5	1029	3	US-08-911-853-1	Sequence 1, Appl
35	34.6	1.5	1029	4	US-09-479-409-1	Sequence 1, Appl
36	34.6	1.5	1029	4	US-09-479-453-1	Sequence 1, Appl
37	34.6	1.5	2040	3	US-08-604-789B-1	Sequence 1, Appl
38	34.6	1.5	2040	3	US-08-604-789B-12	Sequence 12, Appl
39	34.6	1.5	2040	4	US-09-312-721A-1	Sequence 1, Appl
40	34.6	1.5	2040	4	US-09-312-721A-12	Sequence 12, Appl
41	34.6	1.5	4377	3	US-08-911-853-28	Sequence 28, Appl
42	34.6	1.5	4377	4	US-09-479-409-28	Sequence 28, Appl
43	34.6	1.5	4377	4	US-09-479-453-28	Sequence 28, Appl
44	34.6	1.5	10348	2	US-08-457-273B-41	Sequence 41, Appl
45	34.6	1.5	10348	3	US-08-556-419-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1  
US-08-446-345-35  
; Sequence 35, Application US/08446345  
; Patent No. 5831009  
; GENERAL INFORMATION:  
; APPLICANT: Ullrich, Axel  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE  
; TITLE OF INVENTION: PHOSPHATASES PTP-D1  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,345  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/234,440  
; FILING DATE: 28-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30742  
; REFERENCE/DOCKET NUMBER: 7683-054  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)790-9090  
; TELEFAX: (212) 869-8864  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4080 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
US-08-446-345-35

Query Match 2.4%; Score 56.8; DB 2; Length 4080;  
Best Local Similarity 46.6%; Pred. No. 3.7e-06;  
Matches 218; Conservative 0; Mismatches 247; Indels 3; Gaps 1;



[illegible]

	Query Match	1.7%	Score 40.8	DB 2	Length 4776
	Best Local Similarity	43.9%	Pred. No. 0.18		
	Matches 174	Conservative 0	Mismatches 222	Indels 0	Gaps 0
QY	1862	TCAAGAGCTCCCGCGGCTGGAGAACTTCTCCAGAGACTTTTGAGCTGCAGAAAGGTGTGT	1921		
Db	797	TCAACCGCGGCGGGCATCAAGGAGGACACCCGCCACGCGTGCTGCGCGGTCAACG	856		
QY	1922	ACCTACCGCTCAACACCTTTCCTCTGGCGGCACATGCAACGGCTCATGCACTACAGCAGG	1981		
Db	857	AGATCGCGCACACCGCGCGGCATAGTGTGGCTGCGCGCACCGGGGTAAATCGSCCTGC	916		
QY	1982	TCCTGGAGCGGCTGTGAAACACCAACCGCGCGGACGCGACTTTCAGGCACTGCCGAG	2041		

Query Match 1.7%; Score 39.6; DB 4; Length 4722;  
Best Local Similarity 47.5%; Pred. No. 0.39;

Matches 154; Conservative 0; Mismatches 164; Indels 6; Gaps 1;  
QY 1800 GAAGCACCTGGCGGCTACCTGTGGAAGACACAGCGAGCCCTGGAGAAATGG 1859  
DB 618 GGAGAAGCTGGCGGCTCTGTGCAAGAATATGCGGAATCTGCTCAGGAGACCGGAATC 677  
QY 1860 AATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTGTG 1919  
DB 678 GCAGAAGCAGATGAAGCTGCTGCAGAAGAAGCAGAGCCAGCTGGTGCAGAGGAGG--- 733  
QY 1920 TTACCTACCGTCAACACCTTCCTCTCGGGCCACTGCACCGGCTCATGCACTACAAGCA 1979  
DB 734 --ACCACCTGGTGGCGAGCAGACAGCAAGGCCATCTGCGCCGCGAGCAAGCTCGAGAGCT 791  
QY 1980 GGTCTGGAGCGGCTGTGCAAAACACCCCGCGAGCCAGCCGACTTCAGGAGTGGCG 2039  
DB 792 GTGCCGGAGCTGCAGCGGCACAAACCGCTCGCTCAAGGAAGAAGTGTGAGCGAGCCG 851  
QY 2040 AGCCGCTTTTGCAGAGATCACGGAGATGTTGGCAGAGCTCCACGGTACGATGATCAAGAT 2099  
DB 852 AGAGGAGGAGGAAGCAAGGAGGTGACGTACACACTTCCAGATGACGCTCAGAGCAT 911  
QY 2100 GGAGAATTTCCAGAGCTGCAGCA 2123  
DB 912 TCAGCTGCAGATGGAGCAGCAAA 935

RESULT 6

US-08-949-155-51  
; Sequence 51, Application US/08949155  
; Patent No. 6271436  
; GENERAL INFORMATION:  
; APPLICANT: Piedrahita, Jorge A  
; APPLICANT: Bazer, Fuller W  
; TITLE OF INVENTION: Compositions and Methods for the  
; Generation of Transgenic Animal Species  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ARNOLD, WHITE AND DURKEE  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: US  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/949,155  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/027,338  
; FILING DATE: 11-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/046,094  
; FILING DATE: 09-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hibler, David W.  
; REGISTRATION NUMBER: 41,071  
; REFERENCE/DOCKET NUMBER: TAMK:177  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4267 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-949-155-51

Query Match

Best Local Similarity 47.2%; Pred. No. 0.82;  
Matches 117; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
QY 1225 GCCGCCCTGGCGCCACCGAGAGAGAGGAGGTGCTTAAGGATAGGACCCAGCAGAGT 1284  
DB 3020 GCCCTCTCTGCGGCCAGGAGCTGATAGAGGAGAGCATGAAGGAGGTGAAGCCCTACCGC 3079  
QY 1285 AAACCTCAGCCCCCGCAGCCCAAGCAGCAGGCTCCCTGACTGGCAGTCCCTCAGCTTTCCGAG 1344  
DB 3080 GAGGAGCTGGAGGCGCAGCTGGGCCCGCTGACCCAGGAGAGCAGCGCGCTGTCCAAG 3139  
QY 1345 CTGTCTGTGAATCGCAGGGGGAGTGGCCCTGCGCAACGTGACCTTGTCTCCCAACCTG 1404  
DB 3140 GAGCTGAGGCGCGCAGCGCCGCTGGCGCCGCACATGGAGGAGCTGCGCAACCGCTTG 3199  
QY 1405 AGCCCCGACACAAAGCAGCGCTCTCCCTTATGATCAGCCCGCTGCTGAATGACACAGCGCTGC 1464  
DB 3200 GTGCTTACCGCAGCGAGGTGCAACAACATGTTGGGCCAGACACCGAGGAGCTGCGGAGC 3259  
QY 1465 CCCGGAC 1472  
DB 3260 GCCTGGC 3267

RESULT 7

US-09-819-964-51  
; Sequence 51, Application US/09819964  
; Patent No. 6369294  
; GENERAL INFORMATION:  
; APPLICANT: Piedrahita, Jorge A  
; APPLICANT: Bazer, Fuller W  
; TITLE OF INVENTION: Compositions and Methods for the  
; Generation of Transgenic Animal Species  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ARNOLD, WHITE AND DURKEE  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: US  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/819,964  
; FILING DATE: 28-Mar-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/949,155  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: US 60/046,094  
; FILING DATE: 09-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hibler, David W.  
; REGISTRATION NUMBER: 41,071  
; REFERENCE/DOCKET NUMBER: TAMK:177  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4267 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 51:  
US-09-819-964-51

Query Match	1.6%; Score 38.4; DB 4; Length 4267;
Best Local Similarity	47.2%; Pred. No. 0.82;
Matches 117; Conservative	0; Mismatches 131; Indels 0; Gaps 0;
QY 1225	GCCGCTCGCGGCCAGGAGGAGGAGGTGGTTAAGCATAGGACCCAGCAGAGT 1284 
Db 3020	GCCTCTCTCGCGGCCAGGAGCTATAGAGGAGGATGAAGAGGTGAAGGCTACCGC 3079 
QY 1285	AAACCTCAGCCCCCGCACCAAGCAGAGTCCTCTGACTGGCAGTGCTCACCCTTTCCGAG 1344 
Db 3080	GAGGAGCTGGAGCGCAGCTGGGCCGCTGACCCAGGAGCAGCGCGCTGTCCAAG 3139 
QY 1345	CTGCTCTGAACCTCGCAGGAGGAGTGGCCCTCGCAACGTGACCTTGTCTCCCAACCTG 1404 
Db 3140	GAGCTGAGCGCGCGAGGCGCCGCTGGGCGCGGACATGGAGGACGTGGCGAACCGCTTG 3199 
QY 1405	AGCCCCGACACCAAGCAGGCTCTCCCTTCATCAGCCCGCTGCTGAATGACCAAGGCTGC 1464 
Db 3200	GTGCTCTACCGCAGGAGTGCACACATGTTGGCCAGACACCAGGAGCTGGGAGC 3259 
QY 1465	CCCGGAGC 1472 
Db 3260	CGCTGGC 3267 
RESULT 8	
5352575-6	
; Patent No. 5352575	
; APPLICANT: PETROVSKIS, ERIK A.; POST, LEONARD E.; TIMMINS, JAMES G.	
; TITLE OF INVENTION: PSEUDORABIES VIRUS PROTEIN	
; NUMBER OF SEQUENCES: 12	
; CURRENT APPLICATION DATA:	
; APPLICATION NUMBER: US/07/513,282	
; FILING DATE: 20-APR-1990	
; PRIOR APPLICATION DATA:	
; APPLICATION NUMBER: 100,817	
; FILING DATE: 29-JUN-1987	
; APPLICATION NUMBER: 886,260	
; FILING DATE: 16-JUL-1986	
; APPLICATION NUMBER: 784,787	
; FILING DATE: 04-OCT-1985	
; APPLICATION NUMBER: 801,799	
; FILING DATE: 26-NOV-1985	
; APPLICATION NUMBER: 844,113	
; FILING DATE: 26-MAR-1986	
; SEQ ID NO:6:	
; LENGTH: 1053	
5352575-6	
Query Match	1.6%; Score 38.2; DB 6; Length 1053;
Best Local Similarity	54.0%; Pred. No. 0.43;
Matches 101; Conservative	0; Mismatches 83; Indels 3; Gaps 1;
QY 1113	GGCCAGAGTCCGCGGAGGAGGAGGAGTTCGCGCGGGAGCGCGGGTCCGA 1172 
Db 564	GTGCGAGCAGAGCTCGGGTCTGCGCCCGCGGAGGAGCGCGTTCACCAACCGCGC 623 
QY 1173	CCGAGC--CCCTGCGCGGAGGAGGAGCCCGCGGTATACAAGCAGCGGAGCGCGC 1229 
Db 624	CCGATGAGCCAGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 683 
QY 1230	CTCGGCGCCAGCAGGAGGAGGAGGAGTGTCTTAAGGATAGGACCCAGCAGAGTAACC 1289 
Db 684	CGAGCGCCGCTCCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 743 
QY 1290	TCAGCCC 1296 
Db 744	GGTGCCC 750 
RESULT 9	
US-08-467-822-19	
Sequence 19, Application US/08467822	

QY 505 CACATTGGACAACACAGCAATCAGATTTCCAGCTCCTAGA 548  
 | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 445 GCGATCGTCTGCGAGTGGGATCACATTTCCACTTCTCGA 488

## RESULT 10

```

US-08-432-697-19
; Sequence 19, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2619 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..36
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
; OTHER INFORMATION: sequence.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 756..759
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
; OTHER INFORMATION: sequence.
US-08-432-697-19

```

Query Match 1.6%; Score 37.6; DB 4; Length 2619;  
 Best Local Similarity 51.8%; Pred. No. I.I.;  
 Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 385 ATTGTGCAATCTGAGATTGGGATTTGATGAAGCCTTGGACAGAGACATTTAGCAAAA 444  
 | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 325 ACTATCCACATCCGCTAGAGGATATGCAAAATAGCCCCGGCGAGTCTTCTTAAAA 384  
 | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 445 AATAATACATACCTAGCAAGACGCACTTAGGACAAAATCGTGGAAATTTCCACCATAC 504  
 | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 385 AATGAGACATTAATAACCCGCGCAAGAGCCATTAGTTGAAAGTGAATAATAAA 444  
 QY 505 CACATTGGACAACACAGCAATCAGATTTCCAGCTCCTAGA 548  
 | | | | | | | | | | | | | | | | | | | | | | | | | |  
 Db 445 GCGATCGTCTGCGAGTGGGATCACATTTCCACTTCTCGA 488

## RESULT 11

```

US-08-466-248-19
; Sequence 19, Application US/08466248
; Patent No. 6258359
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2619 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..36
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
; OTHER INFORMATION: sequence.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 756..759
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
; OTHER INFORMATION: sequence.
US-08-466-248-19

```

Query Match 1.6%; Score 37.6; DB 4; Length 2619;

```
Best Local Similarity 51.8%; Pred. No. 1.1;
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 385 ATTGTGCAATCTGAGATTGGGATTTTGATGAAGCCTTGACAGAGACACTTAGCAAAA 444
Db 325 ACTATCACACTCCGGTAGAGGATAATGGCAATTAGCCCCCGGAGGTCTTTTAAAA 384

QY 445 AATAAATACATACCTCAGCAAGACGACAGGACAAAATCGTGAATTTTACCCATAAC 504
Db 385 AATGAGACATTACTATTAAACGCCGCAAGAACCCATTAGCTTGAAGTGAAATAAA 444

QY 505 CACATTGGCAACACACGACAGATCAGATTTCCACGCTCTAGA 548
Db 445 GCGATCGTCTGTCAGTGGGATCACATTTCCACTTCTCGA 488

RESULT 12
US-09-404-879A-147
; Sequence 147, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 562
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-147

Query Match 1.6%; Score 37.2; DB 4; Length 562;
Best Local Similarity 53.4%; Pred. No. 0.58;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 43 GACACCCAGAGGCATTTGAAGTTCACAAAGAGCTTCCTGGGAAGTGCTGTGGATGCA 102
Db 175 GATCGAGAGCTGGAGTTTGCAATCCAGCCAAATACAACTGGAACACAGCTTTTGATCAG 234

QY 103 GTTGTCAACACCTCAACCTCGTGGAGGTGACTATTTGGCTCGAGTTCTCTGATCAC 162
Db 235 GTGGTAAGACTATCGGCTCCGGGAGGTGTGTTTGGCTCCACATGTGGATAAT 294

QY 163 AAAAAGATCACGGTGTGGTGATCT 188
Db 295 AAAGGATTTCTACCTGGCTGAGCT 320

RESULT 13
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostek, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
```

```
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match 1.6%; Score 37.2; DB 2; Length 44377;
Best Local Similarity 43.8%; Pred. No. 6.9;
Matches 162; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 1729 GGCCGCTCAATGCCCAATACAGAGATTACCAAAAGATCGGGATGTCATGCTGAAGAAC 1788
Db 29515 GACTTCGCGACCGCTTCACCGGAGCGGACGACGTCGCCGCGGACGTCGCTGAAC 29574

QY 1789 ATTCAGGATGAAGCACCTGGGGGTCTACCTGTGGAAGCACAGCGGCTTTGGAGGCC 1848
Db 29575 TCGCTCACCAAGAGTTCGTGGACGCTCCCTCGGGCTGCTCGTCCGGCGCGGCTTC 29634

QY 1849 CTGGAGATGGATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTG 1908
Db 29635 CTGGAGCTGGCAAGACCGACGTCCTCGGACCCGCGGAGCGGATCGCCCGCAACACCCCGG 29694

QY 1909 CAGAAGTGTGTACCTACCTCAACACTTCTCTCGGCGCACTGCACCGGCTCATG 1968
Db 29695 GTGCGTACCGGGCGGTTCACCTCAACAGGCGGCGGACGACACTCGCGCGGCTGCTG 29754

QY 1969 CACTACAAGCAGTCTCGAGCGGCTGTGCAACACACCCCGCGGAGCCAGCCGACTTC 2028
Db 29755 CGGGAATGATGACCTGTTTCGCGCGGCGGTGCTCACCGGCTGCCGCTGCTCACCCAC 29814

QY 2029 AGGACTGCCAGCGCTTTGGCAGAGATCAGGAGATGTTGGCAGAGCTTCCACGGTAGC 2088
Db 29815 GACGTGCGCGCGGCGGACGCTTCGCGACCATCATCAGCAGCGCGGACACCCGGAAG 29874

QY 2089 ATGATCAAGA 2098
Db 29875 CTCGTCCTGA 29884

RESULT 14
US-08-804-198-1
```





Db	618	TCCAGCACCAACAGGTCCTCAGCTCCACAGGTCCAGCATCCACAGTCCAGTCCACCAGG	559
Qy	2042	CCGCTTTGGCAGATACCGGAGATGGTGGCACACTCCACGGTACGATGATCAAGATGG	2101
Db	558	TCCAGCACCAACAGGTCACAGCACCACAGGTCCAGCATCCACAGTCCAGTCCACCAGG	499
Qy	2102	AGATTTCCAGAGCTGCAGGAATC	2127
Db	498	TCCAGCACCAACAGGTTCCAAAACCAAC	473

Search completed: December 3, 2002, 22:36:06  
Job time : 15029.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 18:25:06 ; Search time 77.6831 Seconds  
(without alignments)  
11655.134 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_151\_2501

Perfect score: 2351

Sequence: 1 ccttcaggaactgtgtc.....attgaggagcgaagacga 2351

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PURCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PURCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PURCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PURCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PURCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PURCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PURCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	762	32.4	3094	10	US-09-925-297-220
2	297	12.6	379	10	US-09-960-352-154
3	292	12.4	400	10	US-09-960-352-156
4	248.6	10.6	408	10	US-09-783-590-3620
5	237.4	10.1	500	10	US-09-783-590-3621
6	190.8	8.1	335	10	US-09-783-590-3600
7	168.8	7.2	356	10	US-09-783-590-3575
8	154.6	6.6	426	10	US-09-815-343-1490
9	144	6.1	3984	10	US-09-848-294-1
10	142.8	6.1	2872	10	US-09-906-779-3
11	142.8	6.1	4125	12	US-10-044-090-620
12	133.2	5.7	288	10	US-09-783-590-3542
13	106	4.5	1718	9	US-09-764-868-51
14	105.2	4.5	716	9	US-09-764-868-475
15	81.2	3.5	466	10	US-09-728-445-750
16	64.4	2.7	2686	9	US-09-764-868-48
17	64.4	2.7	3380	10	US-09-799-799-1
18	64	2.7	1091	9	US-09-764-868-316
19	57.6	2.5	408	10	US-09-960-352-6773

20	57.6	2.5	417	12	US-10-044-090-624	Sequence 624, App
21	57.4	2.4	3957	12	US-10-005-467-1	Sequence 1, Appli
22	56.8	2.4	4080	10	US-09-962-436-342	Sequence 342, App
23	50.2	2.1	512	10	US-09-960-253-27	Sequence 27, Appl
24	50.2	2.1	2330	10	US-09-960-253-156	Sequence 156, App
25	50.2	2.1	3044	10	US-09-880-107-3718	Sequence 3718, Ap
26	50.2	2.1	3047	10	US-09-864-864-329	Sequence 329, App
27	50.2	2.1	3115	10	US-09-925-299-123	Sequence 123, App
28	43.4	1.8	3786	10	US-09-815-242-7865	Sequence 7865, Ap
29	42.4	1.8	1730	9	US-09-860-670-78	Sequence 78, Appl
30	42.4	1.8	1730	9	US-09-764-868-52	Sequence 52, Appl
31	42.4	1.8	1730	9	US-09-764-868-476	Sequence 476, App
32	40.8	1.7	2172	10	US-09-815-242-4038	Sequence 4038, Ap
33	40.4	1.7	598	10	US-09-770-149-991	Sequence 991, App
34	40.4	1.7	768	9	US-09-938-842A-812	Sequence 812, App
35	39.6	1.7	4722	10	US-09-962-055-14	Sequence 14, Appl
36	39.6	1.7	4722	12	US-10-023-529-14	Sequence 14, Appl
37	39.6	1.7	4722	12	US-10-023-523-14	Sequence 14, Appl
38	39.4	1.7	420	10	US-09-960-352-9773	Sequence 9773, Ap
39	38.8	1.7	300	10	US-09-998-598-1581	Sequence 1581, Ap
40	38.6	1.6	414	10	US-09-954-456-1461	Sequence 1461, Ap
41	38.4	1.6	1685	9	US-09-954-531-133	Sequence 133, App
42	38.4	1.6	1685	9	US-09-954-531-354	Sequence 354, App
43	38.4	1.6	1685	10	US-09-962-436-261	Sequence 261, App
44	38.4	1.6	2296	10	US-09-822-849A-259	Sequence 259, App
45	37.8	1.6	420	10	US-09-960-352-9532	Sequence 9532, Ap

ALIGNMENTS

RESULT 1  
US-09-925-297-220  
; Sequence 220, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 220  
; LENGTH: 3094  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (4)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-297-220

Query Match 32.4%; Score 762; DB 10; Length 3094;  
Best Local Similarity 100.0%; Pred. No. 2.5e-211;  
Matches 762; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1590	GTGGTTTCAGACAGTGTGAGCAAGAGGACGCCATGCCGGAAGCACTGAAAGTCTCAT	1649
Db	62	GTGGTTTCAGACAGTGTGAGCAAGAGGACGCCATGCCGGAAGCACTGAAAGTCTCAT	121
QY	1650	ATTCCCGAATTTTGACCTTGCACAAATTTTCATATAATTTTCTCAAGAAATTCAGCA	1709
Db	122	ATTCCCGAATTTTGACCTTGCACAAATTTTCATATAATTTTCTCAAGAAATTCAGCA	181
QY	1710	ACGACTTGCCCTGTGGGAAGGCGCTCAAATGCCCAATCAGAGATTACCAAGAAATCGG	1769

```
|||||
Db 182 ACGACTTGCCCTGTGGAGAGCGCTCAAAATGCCAAATCAGAGATTACCAAGAAATCGG 241
|||||
QY 1770 CGATGTCACTGCTCAAGAAATTCAGGCGCATGAAGCACCTCGGGCGCTCACTGTGGAGCA 1829
|||||
Db 242 CGATGTCACTGCTCAAGAAATTCAGGCGCATGAAGCACCTCGGGCGCTCACTGTGGAGCA 301
|||||
QY 1830 CAGCGAGGCCCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 1889
|||||
Db 302 CAGCGAGGCCCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 361
|||||
QY 1890 CTCGAGAGACTTTGAGCTCGAGAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGG 1949
|||||
Db 362 CTCGAGAGACTTTGAGCTCGAGAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGG 421
|||||
QY 1950 GCCACTGCAACCGCTCATGCACTACAGCAGGTCCTGGAGCGGCTGTGAACACCAACCC 2009
|||||
Db 422 GCCACTGCAACCGGCTCATGCACTACAGCAGGTCCTGGAGCGGCTGTGAACACCAACCC 481
|||||
QY 2010 GCGAGCCAGCCGACTTCAGGAGCTCCGAGCGGCTTTGGCAGAGATCACGGAGATGGT 2069
|||||
Db 482 GCGAGCCAGCCGACTTCAGGAGCTCCGAGCGGCTTTGGCAGAGATCACGGAGATGGT 541
|||||
QY 2070 GGCACAGCTCCAGCGTACGATGATCAAGATGAGAAATTTCCAGAACTGCACGAATCAA 2129
|||||
Db 542 GGCACAGCTCCAGCGTACGATGATCAAGATGAGAAATTTCCAGAACTGCACGAATCAA 601
|||||
QY 2130 GAAAGATTGATGGCATTGACAAATCTGTGTTCCGGGAAGGAGTTCATCGTCTGG 2189
|||||
Db 602 GAAAGATTGATGGCATTGACAAATCTGTGTTCCGGGAAGGAGTTCATCGTCTGG 661
|||||
QY 2190 CAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGA 2249
|||||
Db 662 CAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGA 721
|||||
QY 2250 CGTCTGCTATACAGAGCGGGGGCTGAGGGCTCCAATCATGTTTAAAGTCCACGGCA 2309
|||||
Db 722 CGTCTGCTATACAGAGCGGGGGCTGAGGGCTCCAATCATGTTTAAAGTCCACGGCA 781
|||||
QY 2310 GCTCCCGCTCTATGGCATGACGATTCAGAGAGCGGAAGACGA 2351
|||||
Db 782 GCTCCCGCTCTATGGCATGACGATTCAGAGAGCGGAAGACGA 823
|||||

RESULT 2
US-09-960-352-154
; Sequence 154, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 154
; LENGTH: 379
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E1-A9
US-09-960-352-154
Query Match 12.6%; Score 297; DB 10; Length 379;
Best Local Similarity 89.7%; Pred. No. 1.5e-76;
Matches 331; Conservative 0; Mismatches 35; Indels 3; Gaps 1;

QY 1485 TGAGGCGCGGAGGAAGAGATTCCCACTGATAAAGGCTACTTCATAGCTAAAGAACTGTC 1544
|||
Db 11 TGGGGCGCGGAGAAAGAGATTCCCACTGACAAAGCATACTTCATTGCTTAAAGAAGTCTC 70
|||||

RESULT 3
US-09-960-352-156
; Sequence 156, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 156
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E2-A9
US-09-960-352-156
Query Match 12.4%; Score 292; DB 10; Length 400;
Best Local Similarity 87.5%; Pred. No. 4.3e-75;
Matches 343; Conservative 0; Mismatches 45; Indels 4; Gaps 2;

QY 1485 TGAGGCGCGGAGGAAGAGATTCCCACTGATAAAGGCTACTTCATAGCTAAAGAACTGTC 1544
|||
Db 9 TGGGGCGCGGAGAAAGAGATTCCCACTGACAAAGCATACTTCATTGCTTAAAGAAGTCTC 68
|||||
QY 1545 TACCACCGAGCGCAACATATCTGAAGCATCTCGAAGTTATCACTTCGTGTTTCAGAGCAC 1604
|||||
Db 69 CACCACTGAGAGACATATCTGAAGCATCTTGAAGTCATCACTTCGTGTTTCAGAGCAC 128
|||||
QY 1605 AGTGAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAAATTTGA 1664
|||||
Db 129 AGTGAGCAAGAGGAGCTCCATGCCGGAACCTTGAGAAGTCTCATATTTCCCGAAATTTGA 188
|||||
QY 1665 ACCTTTGCACAAATTTCACTAAATTTCTCAAGGAAATTTGAGCAACGACTTGCCTCTGTG 1724
|||||
Db 189 ACCTTTGCACAAATTTCAACATATTTTCTCAAGGACATTTGAGCAACGACTTGCCTCTGTG 248
|||||
QY 1725 GGAAGGCCCTCAAAATGCCCAATCA---GAGATTACCAAGAAATCGGCGATGTCATGCT 1781
|||||
Db 249 GGAAGGCCCTCGAATGCCCATCATGAGAGATTAACCAAGATTCGGAATTCGAGATGTAGTCT 308
|||||
QY 1782 GAAGAACATTTCAGGCGCATGAAGCACCTG-GCGGCTCACCTGTGGAGACACAGCGAGGCCT 1840
|||||
```

```
QY 1545 TACCACCGAGCGAAACATATCTGAAGGATCTCGAAGTTATCACTTCGTGTTTCAGAGCAC 1604
|||||
Db 71 CACCACTGAGAGGACATATCTGAAGGATCTTGAAGTCATCACTTCGTGTTTCAGAGCAC 130
|||||
QY 1605 AGTGAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAAATTTGA 1664
|||||
Db 131 AGTGAGCAAGAGGAGCTCCATGCCGGAACCTTGAAGAGTCTCATATTTCCCGAAATTTGA 190
|||||
QY 1665 ACCTTTGCACAAATTTCACTAAATTTTCTCAAGGAAATTTGAGCAACGACTTGCCTCTGTG 1724
|||||
Db 191 ACCTTTGCACAAATTTCAACAAATTTTCTCAAGGAAATTTGAGCAACGACTTGCCTCTGTG 250
|||||
QY 1725 GGAAGGCCCTCAAAATGCCCAATCA---GAGATTACCAAGAAATCGGCGATGTCATGCT 1781
|||||
Db 251 GGAAGGCCCTCGAATGCCCATCATGAGAGATTAACCAAGAAATTCGAGATGTAATGCT 310
|||||
QY 1782 GAAGAACATTTCAGGCGCATGAAGCACCTGCGGCTCACTGTGGAAGCACAGCGAGGCCTT 1841
|||||
Db 311 GAAGAAGATTTCAGGGGATGAAGCAACTGCGGCCCACTTGTGGAAGCACAGCGAGGCCT 370
|||||
QY 1842 GGAGGCCCT 1850
|||||
Db 371 GGAGGCCCT 379
|||||

RESULT 3
US-09-960-352-156
; Sequence 156, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 156
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E2-A9
US-09-960-352-156
Query Match 12.4%; Score 292; DB 10; Length 400;
Best Local Similarity 87.5%; Pred. No. 4.3e-75;
Matches 343; Conservative 0; Mismatches 45; Indels 4; Gaps 2;

QY 1485 TGAGGCGCGGAGGAAGAGATTCCCACTGATAAAGGCTACTTCATAGCTAAAGAACTGTC 1544
|||
Db 9 TGGGGCGCGGAGAAAGAGATTCCCACTGACAAAGCATACTTCATTGCTTAAAGAAGTCTC 68
|||||
QY 1545 TACCACCGAGCGCAACATATCTGAAGCATCTCGAAGTTATCACTTCGTGTTTCAGAGCAC 1604
|||||
Db 69 CACCACTGAGAGACATATCTGAAGCATCTTGAAGTCATCACTTCGTGTTTCAGAGCAC 128
|||||
QY 1605 AGTGAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAAATTTGA 1664
|||||
Db 129 AGTGAGCAAGAGGAGCTCCATGCCGGAACCTTGAGAAGTCTCATATTTCCCGAAATTTGA 188
|||||
QY 1665 ACCTTTGCACAAATTTCACTAAATTTTCTCAAGGAAATTTGAGCAACGACTTGCCTCTGTG 1724
|||||
Db 189 ACCTTTGCACAAATTTCAACATATTTTCTCAAGGACATTTGAGCAACGACTTGCCTCTGTG 248
|||||
QY 1725 GGAAGGCCCTCAAAATGCCCAATCA---GAGATTACCAAGAAATCGGCGATGTCATGCT 1781
|||||
Db 249 GGAAGGCCCTCGAATGCCCATCATGAGAGATTAACCAAGATTCGGAATTCGAGATGTAGTCT 308
|||||
QY 1782 GAAGAACATTTCAGGCGCATGAAGCACCTG-GCGGCTCACCTGTGGAGACACAGCGAGGCCT 1840
|||||
```

```

Db      309 GAAGAACATTACGGGATGAAGCACTGAGCTGCTCATTGTGGAAGCACAGCTAGGCC 368
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      1841 TGGAGGCCCTGGAGAATGGGAATCAAGAGCTCC 1872
      ||||| ||||| | || ||| || |||
Db      369 TGGAGGCGCTGGACATCGGCACCTCGGCCCC 400

RESULT 4
US-09-783-590-3620
; Sequence 3620, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3620
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (340)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (352)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (356)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (399)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (402)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (406)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3620

Query Match      10.6%; Score 248.6; DB 10; Length 408;
Best Local Similarity 85.0%; Pred. No. 1.9e-62;
Matches 347; Conservative 0; Mismatches 46; Indels 15; Gaps

Qy      1621 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGACCTTTGCACAAATTT 1680
      | | | | |
Db      1 GCACAGCGGAGACACTGAAAAGTCTCATATTCGCCGAATTTTGACCTTTGCACAAATTT 60

Qy      1681 CATACTAATTTTCTCAAGGAAATTGAGCAACGACTTGCCTCTGTGGGAAGCCGCTCAAAAT 1740
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61 CATACTAATTTTNTCAAGGAAATTGAGCAACGACTTGCCTGTGGGAAGCCGCTCAAAAT 120

Qy      1741 GCCCAATCAGAGATTACCAAGAATTCGGCGATGTATGCTCAAGAACAAATTCAGGCGCATG 1800
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (413)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (436)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (450)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (455)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (459)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (461)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (475)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (477)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (487)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (499)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-3621

Query Match 10.1%; Score 237.4; DB 10; Length 500;  
Best Local Similarity 83.3%; Pred. No. 3.9e-59;  
Matches 415; Conservative 0; Mismatches 60; Indels 23; Gaps 13;  
QY 1621 GCCATGCCGGAAGCACTGAAAGTCTCATATCCCGAATTTTGAACCTTTGCACAAATTT 1680  
DB 1 GGCAGAGCGGAAGCACTGAAAGTCTCATATCCCGAATTTTGAACCTTTGCACAAATTT 60  
QY 1681 CATACATAATTTTCAAGAAATTTAGCAACGACTTGCCTGTG-GGGAAGCGCGTCAAA 1739  
DB 61 CATACATAATTTTCAAGAAATTTAGCAACGACTTGCCTGTG-GGGAAGCGCGTCAAA 120  
QY 1740 TGCCCAAAATCAGAGATTACCAAGATTCGGGATGCTGCTGAAGGAACATTCAGGGCA 1798  
DB 121 TGCCCAAAATCAGAGATTACCAAGATTCGGGATGCTGCTGAAGGAACATTCAGGGCA 180  
QY 1799 TGAAGCACCTGGCGG-CTCACCTGT-GGAAGCACAGCG-AGGCCCTTGGAGGCCCTGGAGA 1855  
DB 181 TGAAGCACCTGGCGGTCTCACCTGTGGGAAGCACAGCGAGGCCCTTGGAGGCCCTGGAGN 240  
QY 1856 A--TGAATAACAAGAGTCCCGCGG--CTGGAGAACTTCTGCAGAGACTTT-GAGCTGCA 1910  
DB 241 AATGGNATTCAGAGCTCCCGCGGCTCGAGGAGNCTTCTGCAGAGACTTTGGAGCTGCA 300  
QY 1911 GAGGTGTG-TTACCTACCGCTCAACACCTT--CTTCCTGGCGGCACTGCACCGGCTCAT 1967  
DB 301 GAAGGTGTGTTTACCTACCGCTCAACACCTTCTTCCTNCTCGGGGCCACTGCACCGGTTCT 360  
QY 1968 GCACTACAAG--CAGGTCTCGAGCGGCTGTGCAACACACCCCGCGGAGCCACGCCGA- 2024  
DB 361 GCACTACAGCGAGGTTCTNCTGAGCGGNTNTTGCARAAACACCCCGATCGNGGCCAAG 420  
QY 2025 -----CTTCAGGAGTCCGAGCCGCTTTGGCAGAGATCACGAGATGGTGGCACAGCT 2078  
DB 421 GCCGATTTTCAGGAGTCCGAGCCG-TTTNGCAGNNTTNGAGGTTCTTGGGANANTT 479  
QY 2079 CCACGGTACGATCAAA 2096  
DB 480 CCAGGTTGGTTGTTCAA 497

RESULT 6  
US-09-783-590-3600  
Sequence 3600, Application US/097833590  
Patent No. US20020110850A1  
GENERAL INFORMATION:  
APPLICANT: Haseltine, Patrick J.  
APPLICANT: Li, Haodong  
APPLICANT: Rosen, Craig A.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
FILE REFERENCE: PO-16.2C1  
CURRENT APPLICATION NUMBER: US/09/783,590  
PRIOR FILING DATE: 2000-02-15  
PRIOR APPLICATION NUMBER: 08/420,856  
PRIOR FILING DATE: 1995-04-12  
PRIOR APPLICATION NUMBER: 08/346,731  
PRIOR FILING DATE: 1994-11-21  
NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3600  
LENGTH: 335  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (5)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (29)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (36)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (42)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (72)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (73)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (144)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (157)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (184)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (246)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (254)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (271)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (278)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (298)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (329)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (333)







Qy 795 GGATTTCTGCAAGCTTCTCGAAATCTGTGTGAACATCATGCTTCTTT 846  
 Db 896 CCGATCTTTGCAAAACCTTGTGAAATCCTGTGTGAGCACCATACGTTCTTT 947

RESULT 10  
 US-09-906-779-3  
 : Sequence 3, Application US/09906779  
 : Patent NO. US200206484A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Shi et al.  
 : TITLE OF INVENTION: Human Protein Tyrosine Phosphatase Polynucleotides, Polypeptides  
 : TITLE OF INVENTION: Antibodies  
 : FILE REFERENCE: PT040P1  
 : CURRENT APPLICATION NUMBER: US/09/906,779  
 : CURRENT FILING DATE: 2001-07-18  
 : PRIOR APPLICATION NUMBER: PCT/US01/01563  
 : PRIOR FILING DATE: 2001-01-17  
 : PRIOR APPLICATION NUMBER: 60/176,306  
 : PRIOR FILING DATE: 2000-01-18  
 : NUMBER OF SEQ ID NOS: 7  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 3  
 : LENGTH: 2872  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 US-09-906-779-3

Query Match 6.18; Score 142.8; DB 10; Length 2872;  
 Best Local Similarity 50.3%; Pred. No. 4.2e-31;  
 Matches 463; Conservative 0; Mismatches 442; Indels 15; Gaps 4;

Qy 35 TGCTGGATGACACCCAGGAGGCATTGGAAGTTCCACAAAGAGCTCCTGGGAAGGTGCTGC 94  
 Db 54 TGCTCGAGCGGACCGAAGTGAAGTGGACCTGCGGAACATGCCAAGGCCAGGATTGT 113  
 Qy 95 TGGATGCAGTTTGGCAACACCACTCAACCTCGTGGAGGTGACTATTTTGGCTCGAGTTTC 154  
 Db 114 TTGATCAGATTGTGTACCACCTTGGACCTTGTGGAAACAGATTACTTTGGCTCCAGTTC 173  
 Qy 155 CTGATCAAAAAGATCAGCGTGTGGCTGATCTCTAAACCCATTTGTGAAACAGATTA 214  
 Db 174 TCGACTCTGCCAGGTTGCGCACCTGGCTGGATCATGCCAAACCCATAAAAAGCAGATGA 233  
 Qy 215 GAAGGCCAAAGCAGCTTGTCTTAAGTTTGGTGAAATCTTTTCGCGCTCACCACACAC 274  
 Db 234 AAATGGACCTGCTTATGCTTTACACTTTCAGTTAAATACTATCTTCGAACCAACA 293  
 Qy 275 AACTCCAAGAGAAGAACTCACAAGGTACCTGTTTCGGCTGCAGGTGAAGCAGGCTTGGCTC 334  
 Db 294 ACCTTCGTGAGGAGTTTACAAGGTACCTGTTGTTTGTATACAACTCAGGCATGACATCTTT 353  
 Qy 335 AAGCAGGTTTCACCTGTAATCACACACCGCAGCTCTCTTGATTTCACACATTTGTGCAAT 394  
 Db 354 CTGAAAATTGAAATGGCCCTTATGAAACAGCTTGTGGAATTAGCTGCTCTCTGTCTACAAG 413  
 Qy 395 CTGAGATTGGGATTTTGATGAAGCCTTTGGACA---GAGAGCACTTAGCAAAAATAAAT 451  
 Db 414 CGGAGCTTGGGAGTGGCAGCTTCCAGACACACACAGCAGCTTGTGCTCAGTTTCGGT 473  
 Qy 452 ACATACC---TCAGACAGACGCAATCAGATTCCAGCTCCCTAGAGATTTGCCGTCGGCTAGAGA 508  
 Db 474 TCATTCCAAATCAGACAGAAGCAATTTGATATCTTCCAGAGATGGAACAGTGCA 533  
 Qy 509 TTGGACAACACCAGCAGAATCAGATTTCAGCTCCCTAGAGATTTGCCGTCGGCTAGAGA 568  
 Db 534 GGGAAAGAGCCCTGCCACGCGGAACCTCTCCTATCTGAATAAAAGCGAAGTGGCTGGAAA 593  
 Qy 569 TGTATGGAATCCGCTTGGACCCGCCACGAGGACGGGAAGGCACCAAGATCAATCTGCGCG 628  
 Db 594 TGTATGGGTAGACATGCACGCTTTCAGGGGAAGAGATGCTGTGTAATATTTCTCTTGAC 653  
 Qy 629 TTGCCAACACGGGAATTTCTAGTGTTCAGGGTTTTCACCTAAGATCAATGCCCTTCAACTGGG 688

Db 810 TTGCAGTTGGAGATGACATCGTGTCCGGAAGGCTGCCCTGCTCCTTTGTTACCCCTGGCC 869  
QY 370 CTTCTGATTTACACATTTGCAATCTGAGATTGGGATTTTGA---TGAAGCCTTGGAC 426  
Db 870 TTGCTGGGCTCCTCATGCTCCAGTCAGAGCTCGGAGACTATGACCCAGATGAATGTGGG 929  
QY 427 AGAGACACTTAGCAAAAAATAAATACATACCTCAGCAAG---ACGCACCTAGAGGACAAA 483  
Db 930 AGCGATTACATTAGTAGTGAGTTCCGCTTTGCAACCAACACACTAAAGAACTGGAAGACAAA 989  
QY 484 ATCGTGGAAATTTACCATTAACCAATTTGGCAAAACACACAGCAATTCAGATTTCCAGCTC 543  
Db 990 GTGATCGAGCTGCACAAGAGCCACAGAGAAATGACGCCAGCAGACAGATGCAATTTTC 1049  
QY 544 CTAGAGATTGCCCGTGGCTAGAGATGATGGAATCCGGTTGACCCGCCAAGACAGG 503  
Db 1050 TTGGAANAATGCCAAAAAATATCAATGTATGGGGTATGATTTTACATCATGCTAAGGACTCA 1109  
QY 604 GAAGGCACGAAGATCAATCTGGCGGTTGCCAACACCGGGAATTTCTAGTGTTCAGGGTTTC 663  
Db 1110 GAAGGGGTAGAANAATATGTTAGAGTTTGTGCAAGTGGTCTGTTGATATATCCGACCGG 1169  
QY 664 ACTAAGATCAATGCCCTCAACTGGGCGCAAGGTGCGGAAGCTGAGCTTCAAGAGGAGCGC 723  
Db 1170 CTGCGAATAAACAGATTTCCTGGCCCAAGTTCTAAAGATTTTCATACAAACGGAACAAC 1229  
QY 724 TTTCTCATCAAGCTCGGGCCAGATGCCAATAGTGGTACAGGATACCTTGAATTCCTG 783  
Db 1230 TTTTACATTAAGATCGGGCGGAGAGTTTGAACAAATTTGAAAGCACCATTTGGGTTTAAG 1289  
QY 784 ATGGCAGTCTGGGATTTCTGCAAGTCTTCTGGAATCTGTGTTGAACATCATGCTCTTC 843  
Db 1290 CTGCCAAACCATCGAGCTGCCAAGCGTTTATGGAAGTATGTTGAGCATCATACATTT 1349  
QY 844 TTTAGACT 851  
Db 1350 TTCAGACT 1357

RESULT 12  
US-09-783-590-3542  
; Sequence 3542, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21  
; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3542  
; LENGTH: 288  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (3)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (42)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (72)

; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (75)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (93)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (129)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (142)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (155)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (156)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (166)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (168)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (182)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (253)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (263)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (264)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (267)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (270)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (272)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (282)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-3542  
Query Match 5.7%; Score 133.2; DB 10; Length 288;  
Best Local Similarity 85.8%; Pred. No. 5.8e-29;  
Matches 188; Conservative 0; Mismatches 27; Indels 4; Gaps 4;  
QY 1628 CGGAAGCACTGAAAGTCTCATATTTCCCGAATTTTGAACCTTTGCACAAAATTTCTACTACTA 1687  
Db 8 CGGAAGCACTGAAAGTCTCAAAATTTCCCAATTTTNAACCTTTTGCAAAATTTTAAACTA 67  
QY 1688 ATTTTCTCAAGGAAATTTGAGCAACGACTTGCCTGTGGGAGGCCGCTCAAAATGCCCAAA 1747  
Db 68 ATTTTNAAGGAAATTTAAGCAACGNCNCTTGCCTGTGGAAAGGCCGCTCAAAATGCCCAAA 127  
QY 1748 TCAGAGATTACCAAGAAATCGCGGATGT-CATGCTGAAG-AACATTCAGGGCATGAAGCA 1805  
Db 128 TNA-AGATTTACCAANAATTCGGCGATGTNNATGCTGAAGNANCATTCAGGGCATGNGGCA 186  
QY 1806 CTGCGGGC-TCACCTGTGGAAGCACAGCGAGCCTTG 1843  
Db 187 CTGCGGGCTTCACCTGTGGGAAGCACGCGAGGCGCTTG 225  
RESULT 13

US-09-764-868-51  
; Sequence 51, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 51  
; LENGTH: 1718  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-868-51

Query Match 4.5%; Score 106; DB 9; Length 1718;  
Best Local Similarity 64.9%; Pred. No. 1.6e-20;  
Matches 157; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 2110 CAGAACTCAGCAAGCTCAGAAAGATTGATGGCATTGACAATCTTGTGTCGGGA 2169  
DB 13 CAGAACTCAGCGAGCTGCGGGGACCTGGTGGCATAGAGAACCTCATGCTCTGCG 72  
QY 2170 AGGAGCTTCATCGTCTGGGCGAGCTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGC 2229  
DB 73 AGGAGCTTCATCGTCTGGGCGCTTCCAGCTCACCAGAAAGGGCTTCAGCAGAGG 132  
QY 2230 ATGTTCTTCTGTTCAACGAGCTCTGCTGTATACAGAGCGGGGCTGACGGCTTCCAAT 2289  
DB 133 ATGTTCTTCTGTTCAACGAGCTCTGCTGTATACAGAGCGGGGCTTCAGCAGAGG 192  
QY 2290 CAGTTTAAAGTCCAGCGAGCTCCGCTCTATGGCATGACGATTGAGAGCGGAGAC 2349  
DB 193 CACTTCGGATCCGGGCGCTCTCCCTCCCAAGGCATGCTGTTGAAGAAAGTGATAAC 252  
QY 2350 GA 2351  
DB 253 GA 254

RESULT 14  
US-09-764-868-475  
; Sequence 475, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 475  
; LENGTH: 716  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (715)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-868-475

Query Match 4.5%; Score 105.2; DB 9; Length 716;  
Best Local Similarity 64.0%; Pred. No. 1.5e-20;  
Matches 155; Conservative 2; Mismatches 85; Indels 0; Gaps 0;  
QY 2110 CAGAACTCAGCAAGCTCAGAAAGATTGATGGCATTGACAATCTTGTGTCGGGA 2169  
DB 3 CAGAACTCAGCGAGCTGCKWCGGACCTGGTGGCATAGAGAACCTCATGCTCTGCGC 62

QY 2170 AGGAGTTTCATCGTCTGGGCGAGCTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGC 2229  
DB 63 AGGAGTTTCATCGTCTGGGCGCTGCCTTCACAGCTCACCAGAAAGGGCTTCAGCAGAGG 122  
QY 2230 ATGTTCTTCTGTTCAACGAGCTCTGCTGTATACAGAGCGGGGCTGACGGCTTCCAAT 2289  
DB 123 ATGTTCTTCTGTTCAACGAGCTCTGCTGTATACAGAGCGGGGCTTCAGCAGAGG 182  
QY 2290 CAGTTTAAAGTCCAGCGGAGCTCCGCTCTATGGCATGACGATTGAGAGCGGAGAC 2349  
DB 183 CACTTCGGATCCGGGCGCTCTCCCTCCCAAGGCATGCTGTTGAAGAAAGTGATAAC 242  
QY 2350 GA 2351  
DB 243 GA 244

RESULT 15  
US-09-728-445-750  
; Sequence 750, Application US/09728445  
; Patent No. US20020102543A1  
; GENERAL INFORMATION:  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and  
; TITLE OF INVENTION: Animals  
; FILE REFERENCE: LEX-0102-USA  
; CURRENT APPLICATION NUMBER: US/09/728,445  
; CURRENT FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/168,358  
; PRIOR FILING DATE: 1999-12-01  
; NUMBER OF SEQ ID NOS: 891  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 750  
; LENGTH: 466  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(466)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-728-445-750

Query Match 3.5%; Score 81.2; DB 10; Length 466;  
Best Local Similarity 58.3%; Pred. No. 1.1e-13;  
Matches 162; Conservative 0; Mismatches 113; Indels 3; Gaps 1;  
QY 24 CAAAATCCAGATCGTGGATGACACCCAGGAGCATTTGAAGTTCCACAAAGAGCTCCTGG 83  
DB 118 CAACATCTTCTTCTGGATAACACTGTACAGGCTTCAGAGTTTACAAACATGATCAGG 177  
QY 84 GAAGTGCTGCTGGATGACAGTTTGAACACCTCAACCTCGTGGAGAGTGACTATTTGG 143  
DB 178 GCAAGTTCTGTTGGATATAGTCTTCAAGCATCTTGATTTGACTGAGCGAGACTATTTGG 237  
QY 144 CCTCGAGTTTCTCT---GATCACAAGAGATCAGGCTGCTGGATCTCTCTAAACCCAT 200  
DB 238 TTTACAGTTGGCTGACGATTCCACAGATAACCCAGGCTGGATCCAAACAAACCAAT 297  
QY 201 TGTGAACACAGATTAGAAGGCCAACGACCGTTGTTGTAAGTTTGTGGTCAAAATCTTTCC 260  
DB 298 AAGGAAGCAGCTAAGAGAGAGGATCACCCTTACAAATTTGAACCTTTAGAGTCAAAATCTTTGT 357  
QY 261 GCTGACCAACACACTCCAAAGAAAGTCAAGGT 298  
DB 358 AAGTGACCCCAACAAGTTTACAAGAGAGTATACAAGGT 395

Search completed: December 3, 2002, 22:26:06  
Job time : 115.683 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: December 3, 2002, 16:31:10 ; Search time 2531.22 Seconds  
(without alignments)  
15042.359 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_151\_2501  
Perfect score: 2351  
Sequence: 1 ccttcagagaaactcgtgc.....attgagagagcgaagacga 2351

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vrt:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	883.4	37.6	912	14	BQ881649
2	812.6	34.6	926	14	BQ706499
3	781.8	33.3	896	14	BQ672632
4	781	33.2	923	4	BQ950768
5	762	32.4	774	13	BI760610
6	761.2	32.4	804	12	BG747792

7	757.2	32.2	865	10	BE260677
8	755.8	32.1	784	12	BG763918
9	744.8	31.7	1043	14	BQ072025
10	726.2	30.9	901	12	BG475554
11	705.4	30.0	898	12	BE745887
12	704.6	30.0	835	12	BG829192
13	700.2	29.8	902	12	BE910036
14	687.6	29.2	721	9	AU132546
15	682	29.0	988	14	BQ674703
16	679.6	28.9	942	12	BG323704
17	678.2	28.8	891	12	BG767698
18	660	28.1	680	12	BG825738
19	628.4	26.7	909	14	BQ646351
20	608.4	25.9	968	12	BF206873
21	607.6	25.8	934	12	BF314265
22	605.6	25.8	692	12	BG116225
23	598.8	25.5	704	12	BG769615
24	592	25.2	621	12	BG770181
25	591	25.1	861	12	BG764061
26	590.4	25.1	932	12	BF686586
27	577.6	24.6	890	12	BE907778
28	560.6	23.8	788	12	BG750463
29	551.8	23.5	954	12	BF304259
30	551	23.4	560	10	AW411445
31	544	23.1	618	14	BQ331885
32	529.8	22.5	627	10	AV729533
33	526	22.4	859	9	AL537488
34	507	21.6	649	14	BQ745892
35	494.4	21.0	496	14	BM705217
36	491.2	20.9	995	13	BM549907
37	478.4	20.3	939	12	BF316537
38	475.4	20.2	489	9	AL121548
39	468	19.9	736	14	BM963756
40	463	19.7	861	12	BG420356
41	457	19.4	1026	13	BM551766
42	456.4	19.4	1002	12	BG323892
43	454	19.3	454	14	BM728340
44	446.2	19.0	961	14	BQ068709
45	445.4	18.9	708	13	BI696633

ALIGNMENTS

RESULT 1  
BQ881649  
LOCUS BQ881649 912 bp mRNA linear EST 16-AUG-2002  
DEFINITION AGENCOURT\_8728249 NIH\_MGC\_47 Homo sapiens cDNA clone IMAGE:639275  
5', mRNA sequence.  
ACCESSION BQ881649  
VERSION BQ881649.1 GI:22273657  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 912)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>,  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
plate: LLM2533 row: n column: 12  
High quality sequence stop: 684.  
Location/Qualifiers  
1 .912

FEATURES  
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6339275"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter:
GGCAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT      247 a  235 c  252 g  178 t
ORIGIN
Query Match      37.6%; Score 883.4; DB 14; Length 912;
Best Local Similarity 98.7%; Pred. No. 6.e-235;
Matches 901; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 442 AAAATAATATACATCTCAGCAAGACGACTAGAGGACAAATCGTGAATTTCCACCAT 501
Db 1 AAAATAATATACATCTCAGCAAGACGACTAGAGGACAAATCGTGAATTTCCACCAT 60
QY 502 AACCACATTGGCAACACACAGCAATCAGATTTCAGCTCCTAGAGATTGCCGTCGG 561
Db 61 AACCACATTGGCAACACACAGCAATCAGATTTCAGCTCCTAGAGATTGCCGTCGG 120
QY 562 CTAGAGATGATGGAATCCGGTTGCACCCGGCCAAAGGACAGGAGGACGCAAGATCAAT 621
Db 121 CTAGAGATGATGGAATCCGGTTGCACCCGGCCAAAGGACAGGAGGACGCAAGATCAAT 180
QY 622 CTGGCGGTTCGCAACACGGGAATTCAGTGTTCAGGGTTTCACCTAAGATCAATGCCCTTC 681
Db 181 CTGGCGGTTCGCAACACGGGAATTCAGTGTTCAGGGTTTCACCTAAGATCAATGCCCTTC 240
QY 682 AACTGGCCCAAGTCGGGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGG 741
Db 241 AACTGGCCCAAGTCGGGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGG 300
QY 742 CCAGATGCCAATAGTCGGTACCAGGATACCTTTGGAAATTCCTGATGCCAGTCGGGATTC 801
Db 301 CCAGATGCCAATAGTCGGTACCAGGATACCTTTGGAAATTCCTGATGCCAGTCGGGATTC 360
QY 802 TCGAAGTCCTTCTGGAATATCTGTTGACATCATGCTTCTTTAGACTTTTTCGAAGAG 861
Db 361 TCGAAGTCCTTCTGGAATATCTGTTGACATCATGCTTCTTTAGACTTTTTCGAAGAG 420
QY 862 CCACAAACCAAGCCCAAGCCGCTCTCTTTAGCCGGGTCTATCATTTTCGGTTTCAGTGGT 921
Db 421 CCACAAACCAAGCCCAAGCCGCTCTCTTTAGCCGGGTCTATCATTTTCGGTTTCAGTGGT 480
QY 922 CGGACTTCAGAACGAGTTCCTCGACTATGTTTAAAGAGGAGGACATAGAAGGTGCAGTTT 981
Db 481 CGGACTTCAGAACGAGTTCCTCGACTATGTTTAAAGAGGAGGACATAGAAGGTGCAGTTT 540
QY 982 GAAAGGAAGCAGCAGCAAGATTCAATCTATCCGGAGCCTTGCTTCACAGCTACAACTG 1041
Db 541 GAAAGGAAGCAGCAGCAAGATTCAATCTATCCGGAGCCTTGCTTCACAGCTACAACTG 600
QY 1042 AATTTCGGAAGTCTCGAGCAGTCTCAGCAGACGACCGCTTACATTTTGGAGAGGTGCC 1101
Db 601 AATTTCGGAAGTCTCGAGCAGTCTCAGCAGACGACCGCTTACATTTTGGAGAGGTGCC 560
QY 1102 GAATCTCCAGGGGGCCAGAGCTGCCGGCGAGGAAAGGAACCGAAGGTTTTCGCCGGGGAG 1161
Db 661 GAATCTCCAGGGGGCCAGAGCTGCCGGCGAGGAAAGGAACCGAAGGTTTTCGCCGGGGAG 720
QY 1162 CCGGGGTCCACCCCGAGCCCTCGCGGAGGAGAACGCCCGCGGGTTAACAGCAGCGGCAC 1221
Db 721 CCGGGGTCCACCCCGAGCCCTCGCGGAGGAGAACGCCCGCGGGTTAACAGCAGCGGCAC 780
```

```
QY 1222 GGAGCGCGCTCGCGCCGCCACGAGGAGAGGAGGAGGTCGTTAAAGGATAGGACCCAGCAG 1281
Db 781 GGAGCGCGCTTCGCGCGCCGCCACGAGGAGGAGGAGGTCGTTAAAGGATAGGACCCAGCAG 840
QY 1282 AGTAAACCTCAGCCGCCGCCACGAGCAGGCTCCCTGACTGGCAGTCCTCCTTC 1341
Db 841 AGTAAACCTCAGGCGCCGCCACGAGCAGGCTCCCTGACTGGCAGT-CTCACCTTTTC 899
QY 1342 GAGCTGCTGTGA 1354
Db 900 GAGCTGCTGTGA 912

RESULT 2
LOCUS      BQ706499      926 bp      mRNA      linear      EST 16-JUL-2002
DEFINITION AGENCOURT_8474935 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301634
5', mRNA sequence.
ACCESSION  BQ706499
VERSION    BQ706499.1 GI:21845398
KEYWORDS  EST.
SOURCE     human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 926)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Dr. Mark Watson
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM2517 row: n column: 03
          High quality sequence stop: 601.
          Location/Qualifiers
            1..926
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:6301634"
              /clone_lib="NIH_MGC_113"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
              EcoRI; cDNA made by oligo-dT priming. Directionally cloned
              into EcoRI/XhoI sites using the following 5' adaptor:
              GGCAGGAG(G). Library constructed by Ling Hong in the
              laboratory of Gerald M. Rubin (University of California,
              Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
              Superscript II RT (Life Technologies). Note: this is a
              NIH_MGC Library."
BASE COUNT      242 a  247 c  254 g  183 t
ORIGIN
Query Match      34.6%; Score 812.6; DB 14; Length 926;
Best Local Similarity 98.8%; Pred. No. 3.9e-215;
Matches 829; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1470 GACGACCATGAGGATGAGGCGCGGAGGAGATTCCTCAACTGATAAGCGTACTTCAT 1529
Db 1 GACGACCATGAGGATGAGGCGCGGAGGAGATTCCTCAACTGATAAGCGTACTTCAT 60
QY 1530 AGCTAAGGAAGTGTCTACCACCGAGCGAACATATCTGAAGGATCTCGAAGTTTATCATTTC 1589
Db 61 AGCTAAGGAAGTGTCTACCACCGAGCGAACATATCTGAAGGATCTCGAAGTTTATCATTTC 120
QY 1590 GTGGTTTCAGACACAGTGTGACAAAGAGACGCCCATGCGGGAGGACACTGAAAGTCTCAT 1649
Db 1 GTGGTTTCAGACACAGTGTGACAAAGAGACGCCCATGCGGGAGGACACTGAAAGTCTCAT 1649
```

Db	121	GTGGTTTCAGAGACAGTGGAGCAAGAGAGACGCCATCGCGGAGCACTGAAAGTCTCAT	180
QY	1650	ATTCCCGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAAATTCAGCA	1709
Db	181	ATTCCCGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAAATTCAGCA	240
QY	1710	AGCACTTGCCCTGTGGGAAGGCGCTCAATGCCCAAAATCAGAGATTCACCAAGAAATCGG	1769
Db	241	ACGACTTGCCCTGTGGGAAGGCGCTCAATGCCCAAAATCAGAGATTCACCAAGAAATCGG	300
QY	1770	CGATGTCATGCTGAAGAAATTCAGGCGATGAAGCACTTGGCGGCTCACTGTGGAAGCA	1829
Db	301	CGATGTCATGCTGAAGAAATTCAGGCGATGAAGCACTTGGCGGCTCACTGTGGAAGCA	360
QY	1830	CAGCAGGCGCTTGGAGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT	1889
Db	361	CAGCAGGCGCTTGGAGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT	420
QY	1890	CTGCAGAGACTTTGAGCTGCAGAAGTGTGTACCTTACCGCTCAACACCTTCTCTCTGG	1949
Db	421	CTGCAGAGACTTTGAGCTGCAGAAGTGTGTACCTTACCGCTCAACACCTTCTCTCTGG	480
QY	1950	GCACATGCACCGGCTCATCACTACAGCAGCTCCTGGAGCGGCTGTGCAACACCAACCC	2009
Db	481	GCACATGCACCGGCTCATCACTACAGCAGCTCCTGGAGCGGCTGTGCAACACCAACCC	540
QY	2010	GCGAGCCACGCGGACTTCAGGAGCTTCCGAGCGGCTTTGGCAGAGATCACGGAGATGT	2069
Db	541	GCGAGCCACGCGGACTTCAGGAGCTTCCGAGCGGCTTTGGCAGAGATCACGGAGATGT	600
QY	2070	GGCACAGCTCCAGGTCAGATGATCAAGATGAGAAATTTCCAGAAGCTGCAGAACTCAA	2129
Db	601	GGCACAGCTCCAGGTCAGATGATCAAGATGAGAAATTTCCAGAAGCTGCAGAACTCAA	660
QY	2130	GAAAGATTTGATGGCATTCACAACTCTTGGTCCGGAGGAGTTCATCGTCTGG	2189
Db	661	GAAAGATTTGATGGCATTCACAACTCTTGGTCCGGAGGAGTTCATCGTCTGG	720
QY	2190	CAGCCTCAGCAAGCTTCGCGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACGA	2249
Db	721	CAGCCTCAGCAAGCTTCGCGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACGA	780
QY	2250	CGTCTGCTATACACAG-CCGGGGGCTGACGGCCCTCAATCAGTTTAAAGTCCACGGG	2307
Db	781	CGTCTGCTATACACAGCGCGGGGTGGACGCCCTCCAATCAGTTTAAAGTCCACGGG	839
RESULT 3			
BQ672632			
LOCUS			
DEFINITION BQ672632 896 bp mRNA linear EST 15-JUL-2002			
AGENCOURT_8185043 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6257346			
5', mRNA sequence.			
ACCESSION BQ672632			
VERSION BQ672632.1			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 896)			
NIH-MGC http://mgc.nci.nih.gov/.			
National Institutes of Health, Mammalian Gene Collection (MGC)			
Unpublished (1999)			
Contact: Robert Strausberg, Ph.D.			
Email: c9apbs-re@mail.nih.gov			
Tissue Procurement: ATCC			
cDNA Library Preparation: Rubin Laboratory			
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Agencourt Bioscience Corporation			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			
Plate: LLCM2411 row: h column: 19			

FEATURES		High quality sequence stop: 630.	
Source		Location/Qualifiers	
		1. 896	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:6257346"	
		/tissue_lib="NIH_MGC_102"	
		/tissue_type="epidermoid	
		/lab_host="DH10B (phage-resistant)"	
		/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;	
		Site_2: EcoRI; cDNA made by oligo-dT priming.	
		Directionally cloned into EcoRI/XhoI sites using the	
		following 5' adaptor: GGCACGAG(G). Library constructed	
		by Ling Hong in the laboratory of Gerald M. Rubin	
		(University of California, Berkeley) using ZAP-cDNA	
		synthesis kit (Stratagene) and Superscript II RT (Life	
		Technologies) Note: this is a NIH_MGC Library."	
BASE COUNT		230 a 257 c 253 g 156 t	
ORIGIN			
		Query Match 33.3%; Score 781.8; DB 14; Length 896;	
		Best Local Similarity 98.3%; Pred. No. 1.5e-206;	
		Matches 811; Conservative 0; Mismatches 12; Indels 2; Gaps 2;	
QY	1216	GCGAGCGAGCGCCCTCGCGGCCACGAGGAGGAGGTCGTTAAGGATAGGACC	1275
Db	1	GCGAGCGAGCGCCCTCGCGGCCACGAGGAGGAGGTCGTTAAGGATAGGACC	60
QY	1276	CAGCAGAGTAACCTCAGCCCCCGGAGCCCAAGCAGCAGGCTCCCTGACTGGCAGTCTCTAC	1335
Db	61	CAGCAGAGTAACCTCAGCCCCCGGAGCCCAAGCAGCAGGCTCCCTGACTGGCAGTCTCTAC	120
QY	1336	CTTTCCGAGCTGTCTGGAACCTCGCAGGGGGAGTGCGCCCTGCCAACGTCGACCTTGTCT	1395
Db	121	CTTTCCGAGCTGTCTGGAACCTCGCAGGGGGAGTGCGCCCTGCCAACGTCGACCTTGTCT	180
QY	1396	CCCAACTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGAC	1455
Db	181	CCCAACTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGAC	240
QY	1456	CAGGCTTCCCGGCGGAGCAGTGGAGGCGGCGGAGGAGATGAGGCGCGGAGGAGATGAGG	1515
Db	241	CAGGCTTCCCGGCGGAGCAGTGGAGGCGGCGGAGGAGATGAGGCGCGGAGGAGATGAGG	300
QY	1516	AAAGCTTACTTCATAGCTAAGGAAGTGTCTACCAACGAGCGAATATCTGAAGGATCTC	1575
Db	301	AAAGCTTACTTCATAGCTAAGGAAGTGTCTACCAACGAGCGAATATCTGAAGGATCTC	360
QY	1576	GAAGTTATCATTCTCGTGGTTTCAGACACAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAG	1635
Db	361	GAAGTTATCATTCTCGTGGTTTCAGACACAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAG	420
QY	1636	CTGAAAGTCTCATATTTCCGAAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTC	1695
Db	421	CTGAAAGTCTCATATTTCCGAAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTC	480
QY	1696	AAGGAAATGAGCAACGCTTGCCTTGTGGGAGGCGGCTCAATGCCAAATCAGAGAT	1755
Db	481	AAGGAAATGAGCAACGCTTGCCTTGTGGGAGGCGGCTCAATGCCAAATCAGAGAT	540
QY	1756	TACCAAGAAATCGGCGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGCGGCT	1815
Db	541	TACCAAGAAATCGGCGATGTCATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGCGGCT	600
QY	1816	CACCTGTGAAGCAGACGAGGCGCTTGGAGGCGCTGGAGGAGGAGGAGGAGGAGGAGGAGG	1875
Db	601	CACCTGTGAAGCAGACGAGGCGCTTGGAGGCGCTGGAGGAGGAGGAGGAGGAGGAGGAGG	660
QY	1876	CGGCTGGAAGAACTTCGACAGACTTTGAGCTGCAAGGAGTGTGTACCTTACCGCTCAAC	1935
Db	661	CGGCTGGAAGAACTTCGACAGACTTTGAGCTGCAAGGAGTGTGTACCTTACCGCTCAAC	720
QY	1936	ACCTTCTCTCTGCGGCGCACTGCACCGGCTCATGCAACGAGGTCCTTGG-AGCGGCT	1994



```

|||||
Db 721 ACCTTCTCTCTGGCGCACTGGACCGGCTCATGCATACCAGCAAGCGCTGGAAGCGGT 780
|||||
QY 1995 GTCCAAACACCCCGCG-AGCCAGCGGACTTCAGGCACTGCC 2038
|||||
Db 781 GTCCAAACACCCCGCGAAGCGCGCAACTTCAGGGAACGGC 825
|||||

RESULT 4
BO950768 923 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8842056 Lupski_sciatic_nerve Homo sapiens cDNA clone
DEFINITION IMAGE:6204242 5', mRNA sequence.
ACCESSION BO950768
VERSION BO950768.1 GI:22366246
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 923)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13625 row: d column: 03
High quality sequence stop: 585.
FEATURES
Location/Qualifiers
1..923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6204242"
/clone_lib="Lupski_sciatic_nerve"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-GACTAGTTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 238 a 257 c 239 g 189 t
ORIGIN
Query Match 33.2%; Score 781; DB 14; Length 923;
Best Local Similarity 96.0%; Pred. No. 2.6e-20;
Matches 834; Conservative 0; Mismatches 30; Indels 5; Gaps 3;

QY 1298 CGCAGCCAGCAGCGTCCCTGACTGGCGAGTCCTCACCTTCCGAGCTGTCTGTGAAC 1357
|||||
Db 31 CTCATCCCCAACAGCTCCCTGACTGGGAGTCCTCACCTTCCGAGCTGTCTGTGAAC 90
|||||
QY 1358 CGCAGGGGAGTGGCCCTGCCAACGTGACCTTGTCTCCCAACCTGAGCCCGCACCA 1417
|||||
Db 91 CGCAGGGGAGTGGCCCTGCTGCAACGTGACCTTGTCTCCCAACCTGAGCCCGCACCA 150
|||||
QY 1418 ACAGGCTCTCCCTGATACCGCGCTGCTGAATGACGAGCGCTGCCCGGAGCGAG 1477
|||||
Db 151 AGCAGGCTCTCCCTGATCAGCCGCTGCTGAATGACGAGCGCTGCCCGGAGCGAG 210
|||||

```

```

QY 1478 ATGAGGATGAGGCGCGAGGAAGAGATTCCCAACTGATAAAGCTACTTCATAGCTAAGG 1537
|||||
Db 211 ATGAGGATGAGGCGCGAGGAAGAGATTCCCAACTGATAAAGCTACTTCATAGCTAAGG 270
|||||
QY 1538 AAGTGTCTACCAACCAGGAGCAATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTC 1597
|||||
Db 271 AAGTGTCTACCAACCAGGAGCAATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTC 330
|||||
QY 1598 AGAGCACAGTGAGCAAAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTCCCGA 1657
|||||
Db 331 AGAGCACAGTGAGCAAAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTCCCGA 390
|||||
QY 1658 ATTTTGAACCTTTTGACAAAATTTTCACTAATTTTCTCAAGGAAATTGAGCAACGACTTG 1717
|||||
Db 391 ATTTTGAACCTTTTGACAAAATTTTCACTAATTTTCTCAAGGAAATTGAGCAACGACTTG 450
|||||
QY 1718 CCTCTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGATTCGCGCATGCA 1777
|||||
Db 451 CCTCTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGATTCGCGCATGCA 510
|||||
QY 1778 TGCTGAAGAACAATTCAGGCGCATGAAGCACTTGGCGGCTCACTGTGGAAGCACACGGAGG 1837
|||||
Db 511 TGCTGAAGAACAATTCAGGCGCATGAAGCACTTGGCGGCTCACTGTGGAAGCACACGGAGG 570
|||||
QY 1838 CCTTGGAGGCCCTTGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTTCAGAG 1897
|||||
Db 571 CCTTGGAGGCCCTTGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTTCAGAG 630
|||||
QY 1898 ACTTTGAGCTGCGAAGGTGTGTACCTACCGCTCAACACCTTCCTCTGCGGCCACTGCG 1957
|||||
Db 631 ACTTTGAGCTGCGAAGGTGTGTACCTACCGCTCAACACCTTCCTCTGCGGCCACTGCG 690
|||||
QY 1958 ACCGCTCATGCACTACAAGAGGTCTTGAGCGGCTGTGCAAAACACACCGCGCG--AG 2015
|||||
Db 691 ACCGCTCATGCACTACAAGAGGTCTTGAGCGGCTGTGCAAAACACACCGCGCGGAGC 750
|||||
QY 2016 CCACGCCACTTCAGGAGCTGCCGAGCGC-TTTCGGCAGAGATCAC--GGAGATCGTGCG 2072
|||||
Db 751 CAGCGCCGACTTCAGGAGCTGCCGAGCGGCTTTGGCAGAGATCACCGGAAATGGGTGGC 810
|||||
QY 2073 ACAGTCCACGGTACGATGATCAAGATGGAGAAATTCAGAAAGTGCACGAACTCAAGAA 2132
|||||
Db 811 ACAGTCCACGGTACGATGATGAGAAATTCAGAAAGTGCACGAACTCAAGAA 870
|||||
QY 2133 AGATTGATTGGCATTGACAATCTTGTGG 2161
|||||
Db 871 AGATTGATTGGCATTGACAATCTTGTGGG 899
|||||

```

```

RESULT 5
LOCUS BI760610
DEFINITION 603044980F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5185411 5',
mRNA sequence.
ACCESSION BI760610
VERSION BI760610.1 GI:15752188
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 774)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13625 row: d column: 03
High quality sequence stop: 585.
FEATURES
Location/Qualifiers
1..923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6204242"
/clone_lib="Lupski_sciatic_nerve"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-GACTAGTTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 238 a 257 c 239 g 189 t
ORIGIN
Query Match 33.2%; Score 781; DB 14; Length 923;
Best Local Similarity 96.0%; Pred. No. 2.6e-20;
Matches 834; Conservative 0; Mismatches 30; Indels 5; Gaps 3;

QY 1298 CGCAGCCAGCAGCGTCCCTGACTGGCGAGTCCTCACCTTCCGAGCTGTCTGTGAAC 1357
|||||
Db 31 CTCATCCCCAACAGCTCCCTGACTGGGAGTCCTCACCTTCCGAGCTGTCTGTGAAC 90
|||||
QY 1358 CGCAGGGGAGTGGCCCTGCCAACGTGACCTTGTCTCCCAACCTGAGCCCGCACCA 1417
|||||
Db 91 CGCAGGGGAGTGGCCCTGCTGCAACGTGACCTTGTCTCCCAACCTGAGCCCGCACCA 150
|||||
QY 1418 ACAGGCTCTCCCTGATACCGCGCTGCTGAATGACGAGCGCTGCCCGGAGCGAG 1477
|||||
Db 151 AGCAGGCTCTCCCTGATCAGCCGCTGCTGAATGACGAGCGCTGCCCGGAGCGAG 210
|||||

```

found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLAM11462 row: p column: 20  
High quality sequence stop: 774.  
Location/Qualifiers  
i. .774

FEATURES

source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_116"  
/lab\_host="DH10B"  
/note="Organ: pooled colon, kidney, stomach; Vector:  
pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA  
source anonymous pool of 3 colons, age 26 yo male, 49 yo  
female, 71 yo male colon; 46 yo male kidney, and pool of 2  
stomachs, 62 yo male and 70 yo female. Library is  
oligo-dT primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.4 kb,  
insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by C.  
Gruber (Invitrogen). Research Genetics tracking code  
023. Note: this is a NIH\_MGC Library."  
200 a 214 c 226 g 134 t

Query Match 32.4%; Score 762; DB 13; Length 774;  
Best Local Similarity 99.9%; Pred. No. 4.6e-201;  
Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 958 GGAGGACATAAGAGGTGCGAGTTTGAAGGAGCAGCAGCAAGATTCATTCATCCGGAGC 1017  
Db 1 GGAGGACATAAGAGGTGCGAGTTTGAAGGAGCAGCAGCAAGATTCATTCATCCGGAGC 60  
QY 1018 CTTGCTTCACAGCTCAGCAACTGAATTCGGAAGTGTCTGGAGGAGTCTCAGCAGCACC 1077  
Db 61 CTTGCTTCACAGCTCAGCAACTGAATTCGGAAGTGTCTGGAGGAGTCTCAGCAGCACC 120  
QY 1078 AGCCTTACA-TTTGGAGAGGTGCCGAATCTCCAGGGGGCCAGAGTCTCCGGCGAGGAAA 1136  
Db 121 AGCCTTACAGTTTGGAGAGGTGCCGAATCTCCAGGGGGCCAGAGTCTCCGGCGAGGAAA 180  
QY 1137 GGAACCGAAGTTTCCGCCGGGAGCGGGGTGCGACCCGAGCCCTGCGCCGAGGAGAAG 1196  
Db 181 GGAACCGAAGTTTCCGCCGGGAGCGGGGTGCGACCCGAGCCCTGCGCCGAGGAGAAG 240  
QY 1197 CCCCGGGGTAACAGCAGCGCAGCGAGCCGCTCGGCCCCACCGAGGAGGAGGA 1256  
Db 241 CCCCGGGGTAACAGCAGCGCAGCGAGCGGCTCGGCCCCACCGAGGAGGAGGAGGA 300  
QY 1257 GGTCTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCGCCCGCAGCAGCAGGCTC 1316  
Db 301 GGTCTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCGCCCGCAGCAGCAGGCTC 360  
QY 1317 CTTGACTGGCAGTCTCACCCTTCCTGAGTGTCTGTGAATCGCAGGGGGAGTGGCCCC 1376  
Db 361 CTTGACTGGCAGTCTCACCCTTCCTGAGTGTCTGTGAATCGCAGGGGGAGTGGCCCC 420  
QY 1377 TGCCAAGCTGACCTTGTCTCCCAACCTGAGCCCGCAGCAGCAGGCTCTCCCTTGAT 1436  
Db 421 TGCCAAGCTGACCTTGTCTCCCAACCTGAGCCCGCAGCAGCAGGCTCTCCCTTGAT 480  
QY 1437 CAGCCGCTGCTGAATGACCCAGGCTCTGCCCGCGAGCGAGTGAAGGATGAGGCGCCGGAG 1496  
Db 481 CAGCCGCTGCTGAATGACCCAGGCTCTGCCCGCGAGCGAGTGAAGGATGAGGCGCCGGAG 540  
QY 1497 GAGAGATTTCCCACTGATTAAGCGTACTTTCATAGCTTAAGGAAGTGTACACCGAGCG 1556  
Db 541 GAGAGATTTCCCACTGATTAAGCGTACTTTCATAGCTTAAGGAAGTGTACACCGAGCG 600  
QY 1557 AACATATCTCAAGGATCTCAAGTTCATCTGTTGTTTCAGAGCAGCAGTGAGCAAGA 1616  
Db 601 AACATATCTCAAGGATCTCAAGTTCATCTGTTGTTTCAGAGCAGCAGTGAGCAAGA 660

QY 1617 GGAGCCATGCCGGAAGCACTGAAAGTCTCATATATCCCGAATTTGAACCTTTGCACAA 1676  
Db 661 GGAGCCATGCCGGAAGCACTGAAAGTCTCATATATCCCGAATTTGAACCTTTGCACAA 720  
QY 1677 ATTTCATACTAATTTCTCAAGGAATTTGAGCAACGACTTGCCTCTGTGGGAAG 1730  
Db 721 ATTTCATACTAATTTCTCAAGGAATTTGAGCAACGACTTGCCTCTGTGGGAAG 774

RESULT 6  
BG747792  
LOCUS  
DEFINITION  
602705293F1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:4841975 5',  
mRNA sequence.  
ACCESSION  
BG747792  
VERSION  
BG747792.1 GI:14058445  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 804)  
AUTHORS  
NIH-MGC http://mgc.nci.nih.gov/.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM1675 row: j column: 24  
High quality sequence stop: 796.  
Location/Qualifiers  
1. .804  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_43"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pOTB7; Site.1: XhoI; Site.2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library. |"  
BASE COUNT 225 a 191 c 209 g 179 t  
ORIGIN

Query Match 32.4%; Score 761.2; DB 12; Length 804;  
Best Local Similarity 99.1%; Pred. No. 7.9e-201;  
Matches 797; Conservative 0; Mismatches 3; Indels 4; Gaps 3;

QY 337 GGCAGGTTGACGTGTATGACACCAGCGAGCTCTCTTGATTTTCACATTTGTGCAATCT 396  
Db 2 GGCAGGTTGACGTGTATGACACCAGCGAGCTCTCTTGATTTTCACATTTGTGCAATCT 61  
QY 397 GAGATTGGGGATTTTGATGAAGCTTGGACAGAGAGCATTAGCAAAAATAATACATA 456  
Db 62 GAGATTGGGGATTTTGATGAAGCTTGGACAGAGAGCATTAGCAAAAATAATACATA 121  
QY 457 CCTCAGCAACAGCAGCTAGAGGACAAAATCGTGAATTTCCACATAACCACTTTGGACAA 516  
Db 122 CCTCAGCAACAGCAGCTAGAGGACAAAATCGTGAATTTCCACATAACCACTTTGGACAA 181  
QY 517 ACACCAAGCAGAATCAGATTTTCCAGCTCCTAGAGATTCGCCGCTCGGCTAGAGATGTATGA 576  
|||||

Db	182	ACACCAGCAGAATCAGATTTCCTAGCTCCCTAGAGATTGCCCGTCGGCTAGAGATGTATGGA	241
QY	577	ATCCGGTTGACCCGGCCAGGACAGGAGGACGACGAAGATCAATCTGGCGTTGCCAAC	636
Db	242	ATCCGGTTGACCCGGCCAGGACAGGAGGACGACGAAGATCAATCTGGCGTTGCCAAC	301
QY	637	ACGGGAATTCATGTTTACAGGTTTCACTAAGATCAATGCCCTTCAACTGGGCGCAAGGTG	696
Db	302	ACGGGAATTCATGTTTACAGGTTTCACTAAGATCAATGCCCTTCAACTGGGCGCAAGGTG	361
QY	697	CGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGGCCAGATGCCAATAGT	756
Db	362	CGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGGCCAGATGCCAATAGT	421
QY	757	GCCTACAGGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTCTGCAAGTCTTCTGG	816
Db	422	GCCTACAGGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTCTGCAAGTCTTCTGG	481
QY	817	AAATCTGTGTTGAACATCATGCTTCTTTAGACTTTTTTGAAGAGCCCAAAAGCC	876
Db	482	AAATCTGTGTTGAACATCATGCTTCTTTAGACTTTTTTGAAGAGCCCAAAAGCC	541
QY	877	AAGCCGTCCTCTTTAGCCGGGGTTCATCATTTCCGTTTCAGTGGTCCGACTCAGAAGCAG	936
Db	542	AAGCCGTCCTCTTTAGCCGGGGTTCATCATTTCCGTTTCAGTGGTCCGACTCAGAAGCAG	601
QY	937	GTTCCTGACTATGTT--AAAGAAGGAGGACATAAGAAGTGCAGTTTGAAGGAAGCACA	994
Db	602	GTTCCTGACTATGTTTACACAGAGGAGGACATAAGAAGTGCAGTTTGAAGGAAGCACA	661
QY	995	GCAAGATTCATTTATCCGGAGCGCTTTCACAGCCTCAGAACTGAAATTCGGAAGTG	1053
Db	662	G-AAGATTCATTTATCCGGAGCGCTTTCACAGCCTCAGAACTGAAATTCGGAAGTG	720
QY	1054	CTGGAGCTCTCAGCAGACACAGCCTTACATTTTGGAGAGGTCCGAAATCTCCAGG	1113
Db	721	CTGGAGCTCTCAGCAGACACAGCCTTACATTTTGGAGAGGTCCGAAATCTCCAGG	780
QY	1114	GGCCAGAGCTGCCGGCGAGGAAG	1137
Db	781	GGACAGAGCTGCCGGCGAGGAAG	804
RESULT 7			
LOCUS	BE260677	601154188F1 NIH_MGC_19 Homo sapiens	linear EST 26-OCT-2000
DEFINITION		865 bp mRNA clone IMAGE:3510381 5',	
ACCESSION	BE260677	mRNA sequence.	
VERSION	BE260677.1	GI:9132166	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (bases 1 to 865)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-r@mail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov		
	Plate: LLCM194 row: o column: 22		
	High quality sequence stop: 730.		
FEATURES	Location/Qualifiers		
source	1..865		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		

BASE COUNT	207	a	236	c	255	g	167	t
ORIGIN								
Query Match			32.2%					Score 757.2; DB 10; Length 865;
Best Local Similarity			97.3%					Pred. No. 1.1e-199;
Matches	823;	Conservative	0;	Mismatches	18;	Indels	5;	Gaps 5;
QY	569	TGTATGGAATCCGGTTGCACCCCGCCAAAGACAGGAAGGACGACGAAGATCAATCTGGCCG	628					
Db	1	TGTATGGAATCCGGTTGCACCCCGCCAAAGACAGGAAGGACGACGAAGATCAATCTGGCCG	60					
QY	629	TGCCAACACGGGAATCTAGTGTTCAGGTTTCACTAAGATCAATGCCCTTCAACTGGG	688					
Db	61	TTGCCAACACGGGAATCTAGTGTTCAGGTTTCACTAAGATCAATGCCCTTCAACTGGG	120					
QY	689	CCAAGTGGCGGAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGTCCGGCCAGATG	748					
Db	121	CCNAGTGGCGGAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGTCCGGCCAGATG	180					
QY	749	CCAATAGTCGCTACCAAGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTTCGCAAGT	808					
Db	181	CCAATAGTCGCTACCAAGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTCTGCAAGT	240					
QY	809	CCTTCTGGAATAATCTGTTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAAC	868					
Db	241	CCTTCTGGAATAATCTGTTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAAC	300					
QY	869	CAAAGCCCAAGCCGCTCTTTAGCCGGGGTCATCATTTCCGTTTCAGTGGTCCGACTC	928					
Db	301	CAAAGCCCAAGCCGCTCTCTTTAGCCGGGGTCATCATTTCCGTTTCAGTGGTCCGACTC	360					
QY	929	AGAAGCAGGTTCCTCGACTATGTTTAAAGAGGAGGACATAGAAGTTCAGTTTGAAGGA	988					
Db	361	AGAAGCAGGTTCCTCGACTATGTTTAAAGAGGAGGACATAGAAGTTCAGTTTGAAGGA	420					
QY	989	AGCACAGCAAGATTCATTTCTATCCGGAGCCTTGTCTTACAGCCTCAGAACTGAATTCGG	1048					
Db	421	AGCACAGCAAGATTCATTTCTATCCGGAGCCTTGTCTTACAGCCTCAGAACTGAATTCGG	480					
QY	1049	AAGTGTCTGGAGCAGTCTCAGCAGACACAGCCCTTACATTTGGAGAAGTTCGCCAATCTC	1108					
Db	481	AAGTGTCTGGAGCAGTCTCAGCAGACACAGCCCTTACATTTGGAGAAGTTCGCCAATCTC	540					
QY	1109	CAGGGGCCAGAGCTGCCGGCCAGGAAAGAACCCGAAGTTCCTCCGGGGAGCCGGGT	1168					
Db	541	CAGGGGCCAGAGCTGCCGGCCAGGAAAGAACCCGAAGTTCCTCCGGGGAGCCGGGT	600					
QY	1169	CGACCCAGGCGCTGCGCCGAGGAGAGCCCGCGGGTTCACAAAGAGCGGAGCGGAGCCG	1228					
Db	601	CGACCCAGGCGCTGCGCCGAGGAGAGCCCGCGGGTTCACAAAGAGCGGAGCGGAGCCG	660					
QY	1229	CCTCGGCCCCAGGAGGAGGAGGAGTCTTAAAGATAGGACCCAGCAGAGTAAC	1288					
Db	661	CCTCGGCCCCAGGAGGAGGAGGAGTCTG-TAAGGATAGGA-CCAGCAGAGTTAAC	718					
QY	1289	CTCAGCCCCCGCAGCAACACAGGCTCCCTGACT--GGCAGTCTCTCACCTTTCCGAGCTG	1347					
Db	719	CTCAGCCCCCGCAG-CAAGCAGAGGCTCCCTGACTGGGAGCTCTCACCTTTCCGAGCTG	777					
QY	1348	TCGTGAACTCGCAGGGGGAGTGGCCCTTGCCAAAGCTGTGTCCTCCCAACCTGAGC	1407					

```

Db 778 TCTTGACTCGCAGGGGAGTTGGCCCTCGCCACAGTAACC-TGTTTCCCAACCTGGGC 836
QY 1408 CCGAC 1413
  || ||
Db 837 CCGAC 842

RESULT 8
BG763918          784 bp      mRNA      linear      EST 15-MAY-2001
LOCUS             602736888F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862444 5',
DEFINITION        mRNA sequence.
ACCESSION         BG763918
VERSION           BG763918.1 GI:14074571
KEYWORDS          EST.
SOURCE            human.
ORGANISM           Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 784)
AUTHORS            NIH-MGC http://mgc.nci.nih.gov/.
TITLE              National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL            Unpublished (1999)
COMMENT            Contact: Robert Strausberg, Ph.D.
                   Email: cgapbs-remail.nih.gov
                   Tissue Procurement: ATCC/DCMD/DTP
                   cDNA Library Preparation: Ling Hong/Rubin Laboratory
                   cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                   DNA Sequencing by: Incyte Genomics, Inc.
                   Clone distribution: MGC clone distribution information can be
                   found through the I.M.A.G.E. Consortium/LLNL at:
                   http://image.llnl.gov
                   Plate: LLCM1722 row: o column: 21
                   High quality sequence stop: 761.
FEATURES           Location/Qualifiers
                   1..784
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="IMAGE:4862444"
                       /clone_lib="NIH_MGC_49"
                       /tissue.type="melanotic melanoma, high MDR (cell line)"
                       /lab_host="DH10B (phage-resistant)"
                       /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
                       EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                       into EcoRI/XhoI sites using the following 5' adaptor:
                       GCACGAG(G). Size-selected >500bp for average insert size
                       1.8kb. Library constructed by Ling Hong in the laboratory
                       of Gerald M. Rubin (University of California, Berkeley)
                       using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                       II RT (Life Technologies). Note: this is a NIH_MGC
                       Library. |"
BASE COUNT         203 a 223 c 223 g 135 t
ORIGIN
Query Match       32.1%; Score 755.8; DB 12; Length 784;
Best Local Similarity 99.5%; Pred. No. 2.5e-199;
Matches 779; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1021 GCTTCACGCCCTACACAACTGAATTCGGAAGTCTGGAGAGTCTCAGCAGACCCAGC 1080
Db 2 GCTTCACGCCCTACAGAACTGAATTCGGAAGTCTGGAGAGTCTCAGCAGACCCAGC 61

QY 1081 CTTACATTTGGAGAGGTGCCGAATCTCCAGGGGGCCAGAGTCCGCGGAGGAAAGGAA 1140
Db CTTACATTTGGAGAGGTGCCGAATCTCCAGGGGGCCAGAGTCCGCGGAGGAAAGGAA 121

QY 1141 CGAAGGTTTCCGCGGGAGCGGGGTGCGACCCGAGGCCCTCGCGCGGAGGAGAGCCCC 1200
Db CGAAGGTTTCCGCGGGAGCGGGGTGCGACCCGAGGCCCTCGCGCGGAGGAGAGCCCC 181

QY 1201 CGGGGTAAACAGCAGCGGAGCGCGGCTCGGCCGCCACGGAGGAAGGAGGAGGTC 1260
Db CGGGGTAAACAGCAGCGGAGCGCGGCTCGGCCGCCACGGAGGAAGGAGGAGGTC 241

```

```

QY 1261 GTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCAGCAAGCACAGGCTCCCTG 1320
  |||||
Db 242 GTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCAGCAAGCACAGGCTCCCTG 301
  |||||

QY 1321 ACTGGCAGTCTCACCTTTCCGAGCTGTCTGTGAATCGCAGGGGGAGTGGCCCTGCG 1380
  |||||
Db 302 ACTGGCAGTCTCACCTTTCCGAGCTGTCTGTGAATCGCAGGGGGAGTGGCCCTGCG 361
  |||||

QY 1381 AACGTGACCTTGCTCCCAACCTGAGCCCGCAGCACCACCAAGCAGGCTCTCCCTTTGATCAGC 1440
  |||||
Db 362 AACGTGACCTTGCTCCCAACCTGAGCCCGCAGCACCACCAAGCAGGCTCTCCCTTTGATCAGC 421
  |||||

QY 1441 CGCGTCTGAATGACCAAGGCTGCCCGCGGACGAGCATGAGGATGAGGGCCGAGGAAG 1500
  |||||
Db 422 CGCGTCTGAATGACCAAGGCTGCCCGCGGACGAGCATGAGGATGAGGGCCGAGGAAG 481
  |||||

QY 1501 AGATTCCTCAACTGATTAAGCGTACTTCATAGCTAAGGAAGTGTCTACCGAGCGGAACA 1560
  |||||
Db 482 AGATTCCTCAACTGATTAAGCGTACTTCATAGCTAAGGAAGTGTCTACCGAGCGGAACA 541
  |||||

QY 1561 TATCTGAAGGATCTCGAAGTTATCACTTCGTGCTTTCAGAGCAGACGTGACCAAGAGGAC 1620
  |||||
Db 542 TATCTGAAGGATCTCGAAGTTATCACTTCGTGCTTTCAGAGCAGACGTGACCAAGAGGAC 601
  |||||

QY 1621 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTT 1680
  |||||
Db 602 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTT 661
  |||||

QY 1681 CATACTAATTTTCTCAAGGAAA-TTGAGCAACGACTTGCCTGTGGGAAGCGCGCTCAA 1739
  |||||
Db 662 CATACTAATTTTCTCAAGGAAACTTGAGCAACGACTTGCCTGTGGGAAGCGCGCTCAA 721
  |||||

QY 1740 TGCCCAATCAGAGATTACCAAGAATCGCGATGTCATGC-TGAAGAACAATTCAGGGCA 1798
  |||||
Db 722 TGCCCAATCAGAGATTACCAAGAATCGCGATGTCATGCTTTGACGAACAATTCAGGGCA 781
  |||||

QY 1799 TGA 1801
  |||
Db 782 TGA 784

RESULT 9
LOCUS             BQ072025
DEFINITION        AGENCOURT_6859787 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5928511
                   5', mRNA sequence.
ACCESSION         BQ072025
VERSION           BQ072025.1 GI:19901071
KEYWORDS          EST.
SOURCE            human.
ORGANISM           Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 1043)
AUTHORS            NIH-MGC http://mgc.nci.nih.gov/.
TITLE              National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL            Unpublished (1999)
COMMENT            Contact: Robert Strausberg, Ph.D.
                   Email: cgapbs-remail.nih.gov
                   Tissue Procurement: ATCC
                   cDNA Library Preparation: Rubin Laboratory
                   DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
                   Clone distribution: MGC clone distribution information can be
                   found through the I.M.A.G.E. Consortium/LLNL at:
                   http://image.llnl.gov
                   Plate: LLCM2105 row: c column: 08
                   High quality sequence stop: 626.
FEATURES           Location/Qualifiers
                   1..1043
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"

```

```
/clone="IMAGE:5928511"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT      266 a   290 c   294 g   193 t
ORIGIN
Query Match      31.7%; Score 744.8; DB 14; Length 1043;
Best Local Similarity 99.7%; Pred. No. 3.5e-196;
Matches 746; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1591 TGGTTTCAGACACAGTGTGACGAAGAGGAGCGCATCCGGAAGCACTGAAAGTCTCATTA 1650
DB 1 TGGTTTCAGACACAGTGTGACGAAGAGGAGCGCATCCGGAAGCACTGAAAGTCTCATTA 60
QY 1651 TTCCCGAATTTGAACCTTTGCACAAATTTTCATACATAATTTTCTCAAGGAAATTTGAGCAA 1710
DB 61 TTCCCGAATTTGAACCTTTGCACAAATTTTCATACATAATTTTCTCAAGGAAATTTGAGCAA 120
QY 1711 CGACTTGCCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATACCAAGAATCGCG 1770
DB 121 CGACTTGCCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATACCAAGAATCGCG 180
QY 1771 GATGTCTATGCTGAAGAAATTTGAGGCATGAAGCACTCGCGGCTCACCTGTGGGAAGCAC 1830
DB 181 GATGTCTATGCTGAAGAAATTTGAGGCATGAAGCACTCGCGGCTCACCTGTGGGAAGCAC 240
QY 1831 AGCGAGCCCTTGGAGCCCTGGAGAAATGAATCAAGAGTCCCGGCGCTCAACACCTTCCTCGCGG 1950
DB 241 AGCGAGCCCTTGGAGCCCTGGAGAAATGAATCAAGAGTCCCGGCGCTCAACACCTTCCTCGCGG 300
QY 1891 TGCAGAGACTTTCAGCTGCAGAGGTGTGTACCTACCCCTCAACACCTTCCTCGCGG 1950
DB 301 TGCAGAGACTTTCAGCTGCAGAGGTGTGTACCTACCCCTCAACACCTTCCTCGCGG 360
QY 1951 CCACCTGCACCGGCTCATGCACTACAAGCAGGTCTCGAGCGGCTGTGCAAAACACACCGG 2010
DB 361 CCACCTGCACCGGCTCATGCACTACAAGCAGGTCTCGAGCGGCTGTGCAAAACACACCGG 420
QY 2011 CGAGCCACCGGCTTCAGGAGCTCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTG 2070
DB 421 CGAGCCACCGGCTTCAGGAGCTCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTG 480
QY 2071 GCACAGCTCCACGGTACGATGATCAGATGGAGAAATTTCCAGAGCTGCACGAACCTCAAG 2130
DB 481 GCACAGCTCCACGGTACGATGATCAGATGGAGAAATTTCCAGAGCTGCACGAACCTCAAG 540
QY 2131 AAAGATTTGATTGGCAATTCGACAAATCTTGTTGTTCCGGGAAGGAGTTTCATCGCTCTGGGC 2190
DB 541 AAAGATTTGATTGGCAATTCGACAAATCTTGTTGTTCCGGGAAGGAGTTTCATCGCTCTGGGC 600
QY 2191 AGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCCTGTTCACACGAC 2250
DB 601 AGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCCTGTTCACACGAC 660
QY 2251 GTCCTGCTATACAGAGCGGGGCTGACGGGCTCCAATCAGTTTAAAGTCCACGGGCGAG 2310
DB 661 GTCCTGCTATACAGAGCGGGGCTGACGGGCTCCAATCAGTTTAAAGTCCACGGGCGAG 720
QY 2311 CTCGCCCTCTATGGCATGACGATTGAGG 2338
DB 721 CTCGCCCTCTATGGCATGACGATTGAGG 748
```

```
RESULT 10
BG475554
LOCUS      602491544F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4620346 5',
DEFINITION mRNA sequence.
ACCESSION BG475554
VERSION    BG475554.1 GI:13407833
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTF
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI381 row: h column: 11
            High quality sequence stop: 818.
            Location/Qualifiers
                location            1..901
                organism            /organism="Homo sapiens"
                db_xref              /db_xref="taxon:9606"
                clone                /clone="IMAGE:4620346"
                clone_lib            /clone_lib="NIH_MGC_20"
                tissue_type          /tissue_type="melanotic melanoma"
                lab_host             /lab_host="DH10B (phage-resistant)"
                note                 /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCACGAG(G). Size-selected >500bp for average
                insert size 1.8kb. Library constructed by Ling Hong in
                the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 232 a 243 c 253 g 173 t
ORIGIN
Query Match 30.9%; Score 726.2; DB 12; Length 901;
Best Local Similarity 91.3%; Pred. No. 5e-191;
Matches 817; Conservative 0; Mismatches 68; Indels 10; Gaps 4;
QY 1430 CCTTGATCAGCCGCTGCTGAATGACCGGCTGCCCGCGGACGAGATGAGATGAGG 1489
DB 2 CCTTGATCAGCCGCTGCTGAATGACCGGCTGCCCGCGGACGAGATGAGATGAGG 61
QY 1490 GCCGGAGGAAGAGATTCCCACTGATAAAGCGTACTTACCTAGCTAAGGAAGTGTCTACCA 1549
DB 62 GCCGGAGGAAGAGATTCCCACTGATAAAGCGTACTTACCTAGCTAAGGAAGTGTCTACCA 121
QY 1550 CCGAGGCAACATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGACGACAGTGA 1609
DB 122 CCGAGGCAACATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGACGACAGTGA 181
QY 1610 GCAAGAGAGACGCCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAATTTTGAACCTT 1669
DB 182 GCAAGAGAGAGCGCCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAATTTTGAACCTT 241
QY 1670 TGCACAAATTTTCATACATAATTTTCTCAAGGAAATTTGAGCAACGACATTGCCCTGTGGGAAG 1729
DB 242 TGCACAAATTTTCATACATAATTTTCTCAAGGAAATTTGAGCAACGACATTGCCCTGTGGGAAG 301
QY 1730 GCCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGTCATGCTGCTGAAGAAACA 1789
DB 302 GCCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGTCATGCTGCTGAAGAAACA 361

QY 1790 TTCAGGCGATGAAGCAGCTGGCGGCTCACCTGTGGAGCACAGCGAGGCGCTTGAGGSCC 1849  
|||||  
Db 362 TTCAGGCGATGAAGCAGCTGGCGGCTCACCTGTGGAGCACAGCGAGGCGCTTGAGGSCC 421  
QY 1850 TGGAGATGAATCAAGAGCTCCGCGGCTGGAGACTTCTGCAGAGACTTTGAGCTGC 1909  
|||||  
Db 422 TGGAGATGAATCAAGAGCTCCGCGGCTGGAGACTTCTGCAGAGACTTTGAGCTGC 481  
QY 1910 AGAAGTGTGTACCTACCTACCCTCAACACCTTCTCTCGCGGCACCTGCACCGCTCATGC 1969  
|||||  
Db 482 AGAAGTGTGTACCTACCTACCCTCAACACCTTCTCTCGCGGCACCTGCACCGCTCATGC 541  
QY 1970 ACTAAGCAGGCTCTGGAGGCTGTGCAACACCAACCGCGAGCGACCGACTTCA 2029  
|||||  
Db 542 ACTAAGCAGGCTCTGGAGGCTGTGCAACACCAACCGCGAGCGACCGACTTCA 601  
QY 2030 -GGAGCTCGCGAGCGCTTTGGCAGAGATCACGAGATGCTGSCACAGCTCCACGGTAGC 2088  
|||||  
Db 602 GGGAGCTGGCGAGCGCTTTGGCAGAGATCACGAGATGCTGSCACAGCTCCACGGTAGC 661  
QY 2089 ATGATCAAGATGAGAAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATT 2148  
|||||  
Db 662 ATGATCAAGATGAGAAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATT 721  
QY 2149 GACAACTTTGTGTTCGGGAA----GGGAGTTTCATCCGCTCTGGGCGCTCAGCAA--- 2201  
|||||  
Db 722 GACAACTTTGTGTTCGGGAAAGCGAGTTCCATCCGCTCTGGGCGCTTCAAGCAAGC 781  
QY 2202 -GCTCTCGGGAGGGGCTCCAGCGGCATGTTCTTCTGTTTCAAGGAGTCTT-GCTA 2259  
|||||  
Db 782 TCTCGGGGAGGGAGTCTCCAGCAGCGCATGCTTCTTGTGTCACGAGTCTCTGGCTT 841  
QY 2260 TACACGAGCGGGGCTGACGGGCTCCATCAGTTTAAAGTCCACGGCGAGCTCC 2314  
|||||  
Db 842 TACACGAGCGGGGCTGACAGGGCTCCATCCAGTTTAAAGCCCAAGGCGAGTCCC 896  
RESULT 11  
BE745887 898 bp mRNA linear EST 15-SEP-2000  
LOCUS 601573513F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3834699 5',  
DEFINITION mRNA sequence.  
ACCESSION BE745887  
VERSION BE745887  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 898)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@remail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LNCM515 row: i column: 04  
High quality sequence stop: 723.  
Location/Qualifiers  
1..898  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3834699"  
/clone\_lib="NIH\_MGC\_9"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:

ECORI: cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).\*

BASE COUNT 236 a 229 c 269 g 164 t  
ORIGIN

Query Match 30.0%; Score 705.4; DB 12; Length 898;  
Best Local Similarity 96.8%; Pred. No. 3.2e-185;  
Matches 762; Conservative 0; Mismatches 21; Indels 4; Gaps 4;

QY 511 GGACAAACACGACAGAAATCCAGATTCCTAGAGATTGCCCTCGGCTAGAGATG 570  
|||||  
Db 1 GGACAAACACGACAGAAATCCAGATTCCTAGAGATTGCCCTCGGCTAGAGATG 60  
QY 571 TATGGAATCCGGTTGCACCGGCAAGGACAGGAGGACGAGATCAATCTGCGCGTT 630  
|||||  
Db 61 TATGGAATCCGGTTGCACCGGCAAGGACAGGAGGACGAGATCAATCTGCGCGTT 120  
QY 631 GCCAACACGGGAATTCAGTGTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGGCC 690  
|||||  
Db 121 GCCAACACGGGAATTCAGTGTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGGCC 180  
QY 691 AAGGTGCGGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGCCAGTCTGCAAGTCC 750  
|||||  
Db 181 AAGGTGCGGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGCCAGTCTGCAAGTCC 240  
QY 751 AATAGTCCGTACCAGGATACCTTTGGAATTCCTGATGCCAGTGGGATTTCTCAAGTCC 810  
|||||  
Db 241 AATAGTCCGTACCAGGATACCTTTGGAATTCCTGATGCCAGTGGGATTTCTCAAGTCC 300  
QY 811 TTCTGGAATCTGTGTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAACCA 870  
|||||  
Db 301 TTCTGGAATCTGTG-TGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAACCA 359  
QY 871 AAGCCCAAGCGCTCTCTTTAGCGGGGTCTATCTTTTCGCTTCAGTGTTCGAGTCTCAG 930  
|||||  
Db 360 AAGCCCAAGCGCTCTCTTTAGCGGGGTCTATCTTTTCGCTTCAGTGTTCGAGTCTCAG 419  
QY 931 AAGCAGGTTCTCGACTATGTTAA-AGAAGGAGGACATTAAGAAGGTGCAGTTTGAAGAGAA 989  
|||||  
Db 420 AAGCAGGTTCTCGACTATGTTAAACAGAAGGAGGACATAAGAAGGTGCAGTTTGAAGAGAA 479  
QY 990 GCACAGCAAGATTTCATTCTATCCGGAGCCTTGTCTCACACCTTACAGAACTGAATTCGGA 1049  
|||||  
Db 480 GCACAGCAAGATTTCATTCTATCCGGAGCCTTGTCTCACAGCCTTACAGAACTGAATTCGGA 539  
QY 1050 AGTGCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAGAGTCCCGAATCTCC 1109  
|||||  
Db 540 AGTGCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAGAGTCCCGAATCTCC 599  
QY 1110 AGGGGCCAGAGTGTCCGCGAGGAAAGGAAACCGAGTTTCCCGCGGGAGCGGGGTC 1169  
|||||  
Db 600 A-GGGGCCAGAGTGTCCGCGAGGAAAGGAAACCGAGTTTCCCGCGGGAGCGGGGTC 658  
QY 1170 GCACCCGAGCCCTGGCGCGAGGAGAGAGCCCGCGGGTAAACAAGCAGCGGACGAGCCGC 1229  
|||||  
Db 659 GCACCCGAGCCCTGGCGCGAGGAGAGAGCCCGCGGGTAAACAAGCAGCGGACGAGCCGC- 717  
QY 1230 CTCGGGCCCCACGGAGGAGGAGGAGGTCGTTAAGGATAGACCCAGCAGAGTAAACC 1289  
|||||  
Db 718 CTCGGGCCCCACGGAGGAGGAGGAGGAGGCGCCGTTAAGTAGGAGCCCAAGAGAGTAAAC 777  
QY 1290 TCAGCCC 1296  
|||||  
Db 778 TTAGCCC 784  
RESULT 12  
BG829192

LOCUS BG829192 835 bp mRNA linear EST 22-MAY-2001  
DEFINITION 602753246F1 NTH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4906114 5',  
mRNA sequence.  
ACCESSION BG829192  
VERSION BG829192.1 GI:14176766  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 835)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1808 row: k column: 11  
High quality sequence stop: 835.  
Location/Qualifiers  
1. 835  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4906114"  
/clone\_lib="NTH\_MGC\_17"  
/tissue\_type="Thadomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; vector: pOTB7; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dr priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 213 a 231 c 243 g 148 t  
ORIGIN  
Query Match 30.0%; Score 704.6; DB 12; Length 835;  
Best Local Similarity 95.8%; Pred No. 5.1e-185;  
Matches 800; Conservative 0; Mismatches 24; Indels 11; Gaps 7;  
QY 826 GTTGAACATCATGCGCTCTTTAGACTTTTGAAGAGCCCAACCAAGCCCAAGCCCGTC 885  
Db 2 GTTGAACATCATGCGCTCTTTA-ACITTTTGAAGAGCCCAACCAAGCCCAAGCCCGTC 60  
QY 886 CTCCTTTAGCGGGGTCATCATTTCCGTTTCAGTGTGCGACTCAGAAAGCAGGTTCTCGAC 945  
Db 61 CTCCTTTAGCGGGGTCATCATTTCCGTTTCAGTGTGCGACTCAGAAAGCAGGTTCTCGAC 120  
QY 946 TATGTTAAAGAGGAGGACATAGAAGGTGCAGTTTGAAGGAAGCACACGCAAGATTTCAT 1005  
Db 121 TATGTTAAAGAGGAGGACATAGAAGGTGCAGTTTGAAGGAAGCACACGCAAGATTTCAT 180  
QY 1006 TCTATCCGAGCGCTTCTTCACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCT 1065  
Db 181 TCTATCCGAGCGCTTCTTCACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCT 240  
QY 1066 CAGCAGACACAGCCTTACATTGAGAAAGTGCCGAATCTCCAGGGGGCCAGAGCTGC 1125  
Db 241 CAGCAGACACAGCCTTACATTGAGAAAGTGCCTGGAATCTCCAGGGGGCCAGAGCTGC 300  
QY 1126 CGCGAGGAAGAAAGCAACCGAGGTTTCCGCCGGGAGCCGGGTGCGACCCGAGCCCTCGG 1185  
Db 301 CGCGAGGAAGAAAGCAACCGAGGTTTCCGCCGGGAGCCGGGTGCGACCCGAGCCCTCGG 360  
QY 1186 CGGAGGAGAGCCCGCGGGTAAACAGAGCGGAGCGCGCTCGGCGCCCGACGGAG 1245

Db 361 CGGAGGAGAAGCCCGCGGGTAAACAAGAGCGGACGGAGCGCGCTCGGCGCCCGACGGAG 420  
QY 1246 GAAGAGGAGGAGTGGTT-AAGGATAGGACCCAGCAGAGTAACCTTCAGCCCCCGCAGCC 1304  
Db 421 GAAGAGGAGGAGTGGTTCAAGGATAGGACCCAGCAGAGTAACCTTCAGCCCCCGCAGCC 480  
QY 1305 AAGCAGAGCTCCCTGACTGGCAGTCCCTCACCTTTCCGAGCTGTCTGTGAACCTCGCAGGG 1364  
Db 481 AAGCAGAGCTCCCTGACTGGCAGTCCCTCACCTTTCCGAGCTGTCTGTGAACCTCGCAGGG 540  
QY 1365 GGGAGTGGCCCCCTGCCAACGTGACCTTGTCTCCCAACTGAGCCCCCGCAGCACCAGCAGCC 1424  
Db 541 GGGAGTGGCCCCCTGCCAACGTGACCTTGTCTCCAGACTGAGCCCCGAAACCAAGCAGCC 600  
QY 1425 CTTCTCCCTTGATCAGCCCGCTGC-TGAATGACCAAGGCTGCCCGCGGAGCGGAGATGAGG 1483  
Db 601 CTTCTCCCTTGATCAGCCCGCTGTTGAATGACCAAGGCTGCCCGCGGAGCGGAGATGAGG 660  
QY 1484 ATGAGGGCCGGAGGAGAGATATCCCAACTGATAAGCGTACTTCA-TAGCTAAGGAAGTG 1542  
Db 661 ATGAGGGCCGGAGGAGAGATATCCCAACTGATAAGCGTACTTCA-TAGCTAAGGAAGTG 720  
QY 1543 TCTACCACCGAGCGAAC---ATATCTGAAGGATCTCGAAG-TTATCATTCTGTTGTTTCA 1598  
Db 721 TCTACCACCGAGCGAACCATATTTCTGAAGGATCTCGAAGTTTATCATTCTCGGTTTCA 780  
QY 1599 GAGCAGACTGACCAAGAGGAGCGCATGCCGGAAGCAC---TGAAAAGTCTCAT 1650  
Db 781 AAGCAGACTGAAAGCAAGAGGAGCGCATGCCGGAAGCACCTTGAAAAGTCTCAT 835  
RESULT 13  
BE910036  
LOCUS 601498085F1 NTH\_MGC\_70 Homo sapiens cDNA clone IMAGE:3900021 5',  
mRNA sequence.  
DEFINITION BE910036  
ACCESSION BE910036  
VERSION BE910036.1 GI:10406227  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 902)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9498 row: j column: 22  
High quality sequence stop: 750.  
Location/Qualifiers  
1. 902  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3900021"  
/clone\_lib="NTH\_MGC\_70"  
/tissue\_type="epithelioid carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.1 kb. Library constructed by Life  
Technologies."  
BASE COUNT 231 a 248 c 272 g 151 t  
ORIGIN



Query Match		29.8%;	Score 700.2;	DB 12;	Length 902;
Best Local Similarity		95.8%;	Pred. No. 9e-184;		
Matches 752;		Conservative 0;	Mismatches 28;	Indels 5;	Gaps 3;
QY	1041	GAATTCGGAAGTCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGAGAGGTGC	1100		
Db	1	GAATTCGGAGTCTTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGTGC	60		
QY	1101	CGAATCTCCAGGGGCGCAGAGTTCGGCGAGGAAGAACCGGAAGTTCCTCCGCGGGGA	1160		
Db	61	CGAATCTCCA-GGGGCCAGAGTTCGGCGAGGAAGAACCGGAAGTTCCTCCGCGGGGA	119		
QY	1161	GCCGGGTGCGACCCGAGCCCTGCGCCGAGGAGAACCCCGGGTAAACAGACAGCGGA	1220		
Db	120	GCCGGGTGCGACCCGAGCCCTGCGCCGAGGAGAACCCCGGGTAAACAGACAGCGGA	179		
QY	1221	CGGAGCCGCTCGCGCCCGCCAGGAGAAAGAGGAGGTCTGTTAAGGATAGGACCCAGCA	1280		
Db	180	CGGAGCCGCTCGCGCCCGCCAGGAGAAAGAGGAGGTCTGTTAAGGATAGGACCCAGCA	239		
QY	1281	GAGTAAACCTCAGCCCGCCGAGCAAGCAGAGGCTCCCTGACTGGCAGTCTCTACCTTTC	1340		
Db	240	GAGTAAACCTCAGCCCGCCGAGCAAGCAGAGGCTCCCTGACTGGCAGTCTCTACCTTTC	299		
QY	1341	CGAGCTGTCTGTGAATCTGCAGGGGGAGTGGCCCTGCCAACGTGACCTTGTCTCCAA	1400		
Db	300	CGAGCTGTCTGTGAATCTGCAGGGGGAGTGGCCCTGCCAACGTGACCTTGTCTCCAA	359		
QY	1401	CCTGAGCCCGGACACCAAGCAGGCGCTCTCCCTTGATCAGCCCGCTGCTGAATGACCAAGC	1460		
Db	360	CCTGAGCCCGGACACCAAGCAGGCGCTCTCCCTTGATCAGCCCGCTGCTGAATGACCAAGC	419		
QY	1461	CTGCCCGCGGACGAGCATGAGGATGAGGCGCGGAGGAAGAGATTCCTCAACTGATAAAGC	1520		
Db	420	CTGCCCGCGGACGAGCATGAGGATGAGGCGCGGAGGAAGAGATTCCTCAACTGATAAAGC	479		
QY	1521	GTACTTTCATAGCTAAGGAAGTGTCTACCGACGAGCAACATATCTGAGGATCTCGAAGT	1580		
Db	480	GTACTTTCATAGCTAAGGAAGTGTCTACCGACGAGCAACATATCTGAGGATCTCGAAGT	539		
QY	1581	TATCACTTCGTGTTTCAGAGCACAGTGAAGAAAGAGAGCGCCATGCCGG-AAGCACTGA	1639		
Db	540	TATCACTTCGTGTTTCAGAGCACAGTGAAGAAAGAGAGCGCCATGCCGGAAAGCACTGA	599		
QY	1640	AAA---GTCTCATATTCGGAAATTTGAACTTTGCACAAATTTCTACTAATTTTCTCA	1696		
Db	600	AAAAGTCTCATATTTCCCGAAATTTGAAACCTTTTGCACAAATTTCTACTAATTTTCTCA	659		
QY	1697	AGGAAATGAGCAACGACTTGCCCTCTGGGAAGGCGCTCAATGCCCAATCAGAGATT	1756		
Db	660	AGGAAATGAGCAACGACTTGCCCTCTGGGAAGGCGCTCAATGCCCAATCAGAGATT	719		
QY	1757	ACCAAGAATCGCGATGTCAATGCTGAAGAACATTTACGGGCATGAAGACACCTGCGGCTC	1816		
Db	720	ACCAAGAATCGCGATGTCAATGCTGAAGAACATTTACGGGCATGAAGACACCTGCGGCTC	779		
QY	1817	ACCTG 1821			
Db	780	CTGTG 784			
RESULT 14					
AUI32546					
LOCUS					
DEFINITION					
AUI32546 NT2RP4 Homo sapiens cDNA clone NT2RP4000050 5', mRNA					
ACCESSION					
VERSION					
AUI32546.1 GI:10992900					
KEYWORDS					
EST.					
SOURCE					
human.					
ORGANISM					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 721) Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masubo,Y. and Isogai,T.				
TITLE	HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masubo,Y., Isogai,T.)				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.				
FEATURES	Location/Qualifiers				
Source	1..721				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="NT2RP4000050"				
	/clone_lib="NT2RP4"				
	/cell_type="teratocarcinoma"				
	/cell_line="NT2"				
	/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"				
BASE COUNT	183 a	180 c	214 g	141 t	3 others
ORIGIN					
Query Match	29.2%;	Score 687.6;	DB 9;	Length 721;	
Best Local Similarity	98.8%;	Pred. No. 2.6e-180;			
Matches 712;	Conservative 0;	Mismatches 7;	Indels 2;	Gaps 2;	
QY	559	CGGCTAGAGATGTATGGAAATCGGTTGCACCCGGCCAGGACAGGAGGACGAGATC	618		
Db	1	CGGCTAGAGATGTATGGAAATCGGTTGCACCCGGCCAGGACAGGAGGACGAGATC	60		
QY	619	AATCTGCGCGTTCGCCAACACGGGAATCTAGTGTTCAGGGTTTCTACTAAGATCAATGCC	678		
Db	61	AATCTGCGCGTTCGCCAACACGGGAATCTAGTGTTCAGGGTTTCTACTAAGATCAATGCC	120		
QY	679	TTCACCTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAGCGTTCTCATCAAGCTC	738		
Db	121	TTCACCTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAGCGTTCTCATCAAGCTC	180		
QY	739	CGGCCAGATGCCAATAGTCGCTACCAAGATACCTTGGAATTCCTGATGCCAGTCGGAT	798		
Db	181	CGGCCAGATGCCAATAGTCGCTACCAAGATACCTTGGAATTCCTGATGCCAGTCGGAT	240		
QY	799	TTCGCAAGTCTCTTCGAAAATCTGTGTGAACATCATGCTCTTTAGACTTTTGA	858		
Db	241	TTCGCAAGTCTCTTCGAAAATCTGTGTGAACATCATGCTCTTTAGACTTTTGA	300		
QY	859	GAGCCCAACCAAGCCCAAGCCCGTCTTTTAGCCGGGGTCTCATCTTCGGTTTCAGT	918		
Db	301	GAGCCCAACCAAGCCCAAGCCCGTCTTTTAGCCGGGGTCTCATCTTCGGTTTCAGT	360		
QY	919	GGTCGCACTCAGAAGCAGGTTCCTCGACTATGTTAAAGAGGAGGACATAAGAGCTGCAG	978		
Db	361	GGTCGCACTCAGAAGCAGGTTCCTCGACTATGTTAAAGAGGAGGACATAAGAGCTGCAG	420		
QY	979	TTTGAAGGAAGACACAGCAAGATTCATTTCTATCCGAGGCTTTGCTTTCACAGCCTACAGAA	1038		
Db	421	TTTGAAGGAAGACACAGCAAGATTCATTTCTATCCGAGGCTTTGCTTTCACAGCCTACAGAA	480		
QY	1039	CTGAATTCGGAAGTGTGTGGAGAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGGT	1098		
Db	481	CTGAATTCGGAAGTGTGTGGAGAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGGT	540		

REFERENCE 1 (bases 1 to 721)  
AUTHORS Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y. and Isogai,T.  
TITLE HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y., Isogai,T.)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
FEATURES  
Location/Qualifiers  
1..721  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP4000050"  
/clone\_lib="NT2RP4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SPL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"  
BASE COUNT 183 a 180 c 214 g 141 t 3 others  
ORIGIN

Query Match 29.2%; Score 687.6; DB 9; Length 721;  
Best Local Similarity 98.8%; Pred. No. 2.6e-180;  
Matches 712; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 559 CGCTAGAGATGTATGGAATCCGGTTGCGACCCGCCAAGGACAGGGAAGGCACGAAGATC 618  
Db 1 CGGTAGAGATGTATGGAATCCGGTTGCGACCCGCCAAGGACAGGGAAGGCACGAAGATC 60  
QY 619 AATCTGGCGTTGCCAACACGGGAATTCAGTCTTTTCAGGGTTTCTACTAGATCAATGCC 678  
Db 61 AATCTGGCGTTGCCAACACGGGAATTCAGTCTTTTCAGGGTTTCTACTAGATCAATGCC 120  
QY 679 TTCAACTGGCCCAAGTCGCGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTC 738  
Db 121 TTCAACTGGCCCAAGTCGCGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTC 180  
QY 739 CGCCAGATGCCAATAGTCGTACCAAGGATACCTTGGAAATTCCTGATGCCAGTCGGGAT 798  
Db 181 CGCCAGATGCCAATAGTCGTACCAAGGATACCTTGGAAATTCCTGATGCCAGTCGGGAT 240  
QY 799 TTCTGCAAGTCTTCTGGAAAATCTGTGTGAACATCATGCCCTCTTTAGACTTTTGA 858  
Db 241 TTCTGCAAGTCTTCTGGAAAATCTGTGTGAACATCATGCCCTCTTTAGACTTTTGA 300  
QY 859 GAGCCCAACCAAGCCCAAGCCGTCCTTTAGCCGGGGTTCATCTTCGGTTCA 918  
Db 301 GAGCCCAACCAAGCCCAAGCCGTCCTTTAGCCGGGGTTCATCTTCGGTTCA 360  
QY 919 GTTCGGACTCAGAAGCAGGTTCCTCGACTATGTTAAAGAGGAGGACATAGAAGGTGCAG 978  
Db 361 GTTCGGACTCAGAAGCAGGTTCCTCGACTATGTTAAAGAGGAGGACATAGAAGGTGCAG 420  
QY 979 TTTGAAAGGAAGCAGCAGCAAGATTCATCTATCCGGAGCCTTTCCTTACAGCCTACAGAA 1038  
Db 421 TTTGAAAGGAAGCAGCAGCAAGATTCATCTATCCGGAGCCTTTCCTTACAGCCTACAGAA 480  
QY 1039 CTGAATTCGGAAGTCTGTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGT 1098  
Db 481 CTGAATTCGGAAGTCTGTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGT 540



QY 1099 GCCGAATCCAGGGGCGAGAGCTGCCGGCGAGGAAGAAACCGAAGGTTTCCCGCGGG 1158  
|||||  
Db 541 GCCGATCTCCAGGGGCGAGAGCTGCCGGCGAGGAAGAAACCGAAGGTTTCCCGCGGG 600  
QY 1159 GAG-CCGGGGTGCACCCGAGCCTCGCGCGGAGGAGAACCCCGGGGTAAACAAGCAGGC 1217  
|||  
Db 601 GAGCCCGGGGTGCACCCGAGCCTCGCGCGGAGGAGAACCCCGGGGTAAACAAGCAGGC 660  
QY 1218 GGACGGAGCCGCTCGCGCGCCACGAGGAGGAGGAGGAGTTCG-TTAAGGATAGGACCC 1276  
|||||  
Db 661 GGACGGAGCCGCTNGCGCGCCACGAGGAGGAGGAGGANGTTCGTTAAGGATAGGACCC 720  
QY 1277 A 1277  
|  
Db 721 A 721

RESULT 15  
BQ674703  
LOCUS  
DEFINITION BQ674703 988 bp mRNA linear EST 15-JUL-2002  
5', mRNA sequence.  
ACCESSION BQ674703  
VERSION BQ674703.1 GI:21785537  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 988)  
AUTHORS NIH-MGC <http://mgc.mci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCN2405 row: p column: 02  
High quality sequence stop: 685.  
Location/Qualifiers  
1..988  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6255217"  
/clone\_lib="NIH\_MGC\_102"  
/tissue\_type="epidermoid carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: salivary gland; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed  
by Ling Hong in the laboratory of Gerald M. Rubin  
(University of California, Berkeley) using ZAP-cDNA  
Synthesis Kit (Stratagene) and Superscript II RT (Life  
Technologies). Note: this is a NIH\_MGC Library."  
BASE COUNT 216 a 306 c 291 g 161 t 14 others  
ORIGIN

Query Match 29.0%; Score 682; DB 14; Length 988;  
Best Local Similarity 99.6%; Pred. No. 1.2e-178;  
Matches 693; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1657 AATTGTGAACCTTGCACAAATTCATACATAATTTCTCAAGGAATTCAGCAACGACTT 1716  
|||||  
Db 1 AATTTTGAACCTTGCACAAATTCATACATAATTTCTCAAGGAATTCAGCAACGACTT 60  
QY 1717 GCCCTGTGGGAAGCGCGCTCAAAATGCCCAATCAGAGATTACCAAGAATCGCGGATGTC 1776  
|||||

Search completed: December 3, 2002, 21:25:01  
Job time : 2576.22 secs

Db 61 GCCCTGTGGGAAGCGCGCTCAAAATGCCCAATCAGAGATTACCAAGAATCGCGGATGTC 120  
QY 1777 ATGCTGGAAGAACATTCAGGGCATGAGCACCCTGGCGGCTCACCTGTGGAAGCACAGCGAG 1836  
|||||  
Db 121 ATGCTGAAGAACATTCAGGGCATGAGCACCCTGGCGGCTCACCTGTGGAAGCACAGCGAG 180  
QY 1837 GCCTTTGGAGGCCCTGGAGAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTTCGACA 1896  
|||||  
Db 181 GCCTTTGGAGGCCCTGGAGAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTTCGACA 240  
QY 1897 GACTTTGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGGGGCACTG 1956  
|||||  
Db 241 GACTTTGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGGGGCACTG 300  
QY 1957 CACCGGCTCATGCACACTACAAGCAGGTCTCTGGAGCGGCTGTGCAACACACCCGCGGAGC 2016  
|||||  
Db 301 CACCGGCTCATGCACACTACAAGCAGGTCTCTGGAGCGGCTGTGCAACACACCCGCGGAGC 360  
QY 2017 CACCGCGACTTCAGGACTGCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTGGCACAG 2076  
|||||  
Db 361 CACCGCGACTTCAGGACTGCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTGGCACAG 420  
QY 2077 CTCCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGCAGAACTCAAGAAAGAT 2136  
|||||  
Db 421 CTCCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGCAGAACTCAAGAAAGAT 480  
QY 2137 TTGATTGGCATTGACAATCTTGTGGTTCCGGGAAGGGAGTTTCATCCGTCTGGGCAGCCTC 2196  
|||||  
Db 481 TTGATTGGCATTGACAATCTTGTGGTTCCGGGAAGGGAGTTTCATCCGTCTGGGCAGCCTC 540  
QY 2197 AGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTCCTG 2256  
|||||  
Db 541 AGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTCCTG 600  
QY 2257 CTATACACGAG-CCGGGGGCTGACGGCCTCCCAATCAGTTTAAAGTCCACGGGCGAGCTCC 2315  
|||||  
Db 601 CTATACACGAGNCCGGGGGCTGACGGCCTCCCAATCAGTTTAAAGTCCACGGGCGAGCTCCC 660  
QY 2316 GCTCTATGCGATGAGGATGAGGAGAGCGCAAGACGA 2351  
|||||  
Db 661 GCTCTATGCGATGAGGATGAGGAGAGCGCAAGACGA 696

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: December 3, 2002, 15:35:00 ; Search time 496.761 Seconds  
(without alignments)  
16520.996 Million cell updates/sec  
Title: US-09-555-342B-1\_COPY\_151\_432  
Perfect score: 282  
Sequence: 1 ccttcagaaactgtgtc.....ctgaccacacacactccaa 282

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pi.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	282	100.0	3442	9	AB008430	AB008430 Homo sapi
2	107	37.9	96183	9	AL161896	AL161896 Homo sapi
3	107	37.9	152053	2	AC022669	AC022669 Homo sapi
4	97.8	34.7	2605	9	BC021301	BC021301 Homo sapi
5	97.8	34.7	3997	9	AB018336	AB018336 Homo sapi
6	92.2	32.7	3719	10	BC009153	BC009153 Mus muscu
7	75	26.6	149417	2	AC118779	AC118779 Rattus no
8	73.4	26.0	221502	2	AC122885	AC122885 Mus muscu
9	71.6	25.4	2882	10	AF106702	AF106702 Mus muscu
10	70.8	25.1	170916	9	AL136300	AL136300 Human DNA
11	70.4	25.0	3451	10	AB032366	AB032366 Mus muscu
12	68.4	24.3	3619	9	BC010674	BC010674 Homo sapi
13	68.4	24.3	3643	9	HUMPTYPH	M68941 Human prote
14	65.4	23.2	152053	2	AC022669	AC022669 Homo sapi
15	65.2	23.1	2308	9	AB032179	AB032179 Homo sapi
16	65.2	23.1	3313	6	AX050013	AX050013 Sequence
17	65.2	23.1	3770	9	AF153418	AF153418 Homo sapi
18	65.2	23.1	5582	9	AF153416	AF153416 Homo sapi
19	62.4	22.1	62537	2	AC101100	AC101100 Mus muscu
20	61.4	21.8	14917	6	AX375624	AX375624 Sequence
21	60.6	21.5	3106	6	AX050012	AX050012 Sequence
22	60	21.3	5097	5	AY124488	AY124488 Danio rer
23	59.8	21.2	224310	2	AC094777	AC094777 Rattus no
24	59.4	21.1	2595	9	AF156225	AF156225 Homo sapi
25	58.2	20.6	2964	10	AF044312	AF044312 Mus muscu
26	57.8	20.5	621	6	AX430372	AX430372 Sequence
27	57.8	20.5	2835	9	BC007796	BC007796 Homo sapi
28	57.8	20.5	2867	9	HUMELIA	M14993 Human struc
29	57.8	20.5	3043	9	HUMEMP41	M61733 Homo sapien
30	57	20.2	2717	10	AF177146	AF177146 Mus muscu
31	57	20.2	3661	10	AB032828	AB032828 Rattus no
32	57	20.2	4051	10	AF152247	AF152247 Mus muscu
33	57	20.2	4543	10	AB032827	AB032827 Rattus no
34	56.8	20.1	2625	9	AY049789	AY049789 Homo sapi
35	56.8	20.1	3534	9	BC013885	BC013885 Homo sapi
36	56.8	20.1	6263	9	AB002336	AB002336 Human mRN
37	56.6	20.1	2970	9	AK000388	AK000388 Homo sapi
38	55.2	19.6	5855	10	AB019257	AB019257 Rattus no
39	55.2	19.6	6102	10	AB019256	AB019256 Rattus no
40	55	19.5	1784	9	AK094281	AK094281 Homo sapi
41	54.6	19.4	3984	9	HUMCAP	M64572 Human prote
42	54.6	19.4	5942	10	MUS41SP	L00919 Mus musculu
43	54	19.1	3166	9	AK023019	AK023019 Homo sapi
44	53.4	18.9	2758	5	XELCSK	M20621 X.laevlis cy
45	52.8	18.7	3872	10	BC034751	BC034751 Mus muscu

ALIGNMENTS

RESULT 1  
AB008430  
LOCUS AB008430 3442 bp mRNA linear PRI 13-FEB-1999  
DEFINITION Homo sapiens mRNA for CDEP, complete cds.  
ACCESSION AB008430  
VERSION AB008430.1 GI:2766164  
KEYWORDS CDEP.  
SOURCE Homo sapiens embryo cartilage chondrocyte cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Koyano,Y., Kawamoto,T., Shen,M., Yan,W., Noshiro,M., Fujii,K. and Kato,Y.  
TITLE Molecular cloning and characterization of CDEP, a novel human

protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors

Biochem. Biophys. Res. Commun. 241 (2), 369-375 (1997)

98086358

2 (bases 1 to 3442)

Koyano,Y., Kawamoto,T. and Kato,Y.

Direct Submission

Submitted (22-OCT-1997) Takeshi Kawamoto, Hiroshima University

School of Dentistry, Department of Biochemistry; 1-2-3 Kasumi

Minami-ku, Hiroshima, Hiroshima 734, Japan

(E-mail:tkawano@ipc.hiroshima-u.ac.jp. Tel:082-257-5688,

Fax:082-257-5629)

FEATURES

Source

Location/Qualifiers

1..3442

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_type="chondrocyte"

/tissue\_type="cartilage"

/dev\_stage="embryo"

49..3186

/function="Rho Guanine Nucleotide Exchange Factor"

/note="Band 4.1 superfamily"

/codon\_start=1

/product="CDEP"

/protein\_id="BAA24267.1"

/db\_xref="GI:2766165"

/translation="MGEIQRTPGSRIGAPENSIGSTLGRGQKPPPTSGKLVSIKI  
QMDDTQAFVQVQAPKQVLDLDAVCHNLNLVVEGDFGLQEPDHPKRTITWLDLKPVI  
KOIRDPKVVQVYVFFPPDHTQLEELRYFALQVQDLAQGLRCLTNCDSALLI  
SHVQSEIGDDEALDRHLAKNKYIPQDDALEDKIVFHHNHIGQTPAESDFQLLEI  
ARRLMEYGLRHPKADREGTKINLAVANTGILVFOGFTKINAFNAKVRKLSFKRKR  
LTKLPDANSAYQDTLEFLMASRDFCKSEFKICVEHFAFRLEFPKPKPKPVLFSRG  
SFLRSGTQKQVLYVEGKHKVQFPERKHSKIHISIRLSAQPTLSEVLEQSQOS  
TSLTGEGAESGGSCRGKPEKVSAGEPGSHSPAPRRSPAGNKQADGAASAPTEE  
EEVYKTRDQSKQPPQPGTSLGSHLSVNSQGGVAPANTVLSPLNSPDTKQ  
ASPLSLNDQACPTDEDEGRKRFTDKAYIAKEVSTTERTYTLKDLVITISWF  
QSTVKEDAMPALAKLIFPNPEPLHKFTNFKIEIQRALWEGRSNAQIRDYORIG  
DVMKLNQGMKLAHLWKSLEALENGIKSSRLENFCRDFLOKVCYLPNLTFL  
LRLRLHYKQVLRICKHHPHSHADPRDCRAALAEITEMVAOLHGMIMKRNPKL  
HELKDLGIDNLYPGREFRLGSLKSGKGLQORMPFLFNVDLLTYSGLTRASNQ  
FKVHGOLPIGMTIIESEDEWGVPHCLTLRGQRQSLIIVAASSRSEMERKVEDIQNAID  
LAEKSSPAPELSPDNKSPDEATAADQESDLSASRTSLERQAPHRGNTMVHP  
CWHRTSVSWDFSLAVNQSLNLRKFNKSNQWKLWVVTNFCLEFFYKSHQDNHP  
LASPLGLSLTIPSESENIQKDYVFKLHFKSHVYVFAESEYTPFERMWVIRSATSS  
ASRPHVLSKESLVI"

3442

/note="50 a nucleotides"

BASE COUNT 864 a 952 c 927 g 699 t

polya\_site

BASE COUNT

ORIGIN

Query Match 100.0%; Score 282; DB 9; Length 3442;

Best Local Similarity 100.0%; Pred. No. 1.4e-77;

Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGAAAACCTGTGCCATCAAAATCCAGATCCTGGATGACACCCAGGAGGCATT 60

Db 151 CCTTCAGAAAACCTGTGCCATCAAAATCCAGATCCTGGATGACACCCAGGAGGCATT 210

QY 61 GAAGTTCCCAAGAGCTCCTGGGAAGGTGCTGCTGGATGCAGTTTGCAACCACTCAAC 120

Db 211 GAAGTTCCCAAGAGCTCCTGGGAAGGTGCTGCTGGATGCAGTTTGCAACCACTCAAC 270

QY 121 CTCGTGGAAGGTGACTATTTCGCCTCGAGTTTCCTGATGACAAAAAGATCAGGTGCG 180

Db 271 CTCGTGGAAGGTGACTATTTCGCCTCGAGTTTCCTGATGACAAAAAGATCAGGTGCG 330

QY 181 CTGATCTCTTAAACCCATTGTGAACAGATTAGAGGCCAACGACGTTGTTGTTAAG 240

Db 331 CTGATCTCTTAAACCCATTGTGAACAGATTAGAGGCCAACGACGTTGTTGTTAAG 390

QY 241 TTGTGGTGAATTCCTTCGCGCTGACACACACAACTCCAA 282

|||||

Db 391 TTGTGGTGAATTCCTTCGCGCTGACACACAACTCCAA 432

RESULT 2

AL161896

LOCUS

DEFINITION

complete sequence.

ACCESSION

AL161896

VERSION

AL161896.16

GI:12330752

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 96183)

Phillimore,B.

Direct Submission

Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jan 22, 2001 this sequence version replaced gi:12329254.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated repeat sequence elements. Where the sequence is

ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep

This sequence

was generated from part of bacterial clone contigs of human

chromosome 13, constructed by the Sanger Centre Chromosome 13

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr13

Rpl1-261P24 is from the library RPCI-11.1 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone

Rpl1-261P24. It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true right end of clone Rpl1-261P24 is at 96183 in this

sequence. The true left end of clone Rpl1-31K22 is at 59428 in this

sequence. The true right end of clone Rpl1-573N10 is at 100 in this

sequence.

FEATURES

Location/Qualifiers

1..96183

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="13"

/clone="Rpl1-261P24"

/clone\_lib="RPCI-11.1"

1301..1335

/note="SN5 repeat: matches 34..68 of consensus"

1432..1506

/note="MER20 repeat: matches 1..75 of consensus"

3518..3854

/note="L1MB8 repeat: matches 5837..6173 of consensus"

5528..5599

/note="12 copies 6 mer tatatta 69% conserved"

5533..5604

/note="36 copies 2 mer at 69% conserved"

5766..6068

/note="AluSp repeat: matches 7..303 of consensus"

6949..7264

repeat\_region /note="AluJo repeat: matches 1. .307 of consensus"  
repeat\_region 7380. .7501  
repeat\_region /note="MIR repeat: matches 65. .191 of consensus"  
8192. .8407  
repeat\_region /note="L2 repeat: matches 2064. .2289 of consensus"  
8475. .8506  
repeat\_region /note="16 copies 2 mer tt 100% conserved"  
8929. .9103  
repeat\_region /note="MER33 repeat: matches 4. .171 of consensus"  
9104. .9421  
repeat\_region /note="AluJb repeat: matches 1. .306 of consensus"  
9422. .9450  
repeat\_region /note="MER33 repeat: matches 171. .216 of consensus"  
9491. .9595  
repeat\_region /note="MER33 repeat: matches 142. .225 of consensus"  
9596. .9909  
repeat\_region /note="AluY repeat: matches 1. .311 of consensus"  
9910. .9997  
repeat\_region /note="MER33 repeat: matches 225. .319 of consensus"  
11665. .11763  
repeat\_region /note="MIR repeat: matches 47. .149 of consensus"  
11991. .12022  
repeat\_region /note="8 copies 4 mer tgtg 90% conserved"  
12274. .12558  
repeat\_region /note="AluX repeat: matches 20. .311 of consensus"  
12577. .12879  
repeat\_region /note="AluY repeat: matches 1. .304 of consensus"  
13406. .13549  
repeat\_region /note="MLTIC repeat: matches 1. .146 of consensus"  
13554. .13595  
repeat\_region /note="16 copies 2 mer ca 96% conserved"  
13682. .13715  
repeat\_region /note="17 copies 2 mer ac 97% conserved"  
13684. .13715  
repeat\_region /note="8 copies 4 mer acac 100% conserved"  
13752. .13779  
repeat\_region /note="14 copies 2 mer ca 96% conserved"  
13784. .14048  
repeat\_region /note="MLTIC repeat: matches 189. .465 of consensus"  
14716. .15018  
repeat\_region /note="AluSc repeat: matches 1. .303 of consensus"  
15567. .15842  
repeat\_region /note="AluJo repeat: matches 3. .305 of consensus"  
16752. .16916  
repeat\_region /note="MIR repeat: matches 63. .227 of consensus"  
17700. .17838  
repeat\_region /note="AluJo/FRAM repeat: matches 163. .301 of consensus"  
18111. .18314  
repeat\_region /note="MLTIA2 repeat: matches 157. .359 of consensus"  
18317. .18428  
repeat\_region /note="28 copies 4 mer ctcc 71% conserved"  
18319. .18422  
repeat\_region /note="52 copies 2 mer ct 63% conserved"  
18322. .18423  
repeat\_region /note="17 copies 6 mer tctctc 63% conserved"  
18510. .18797  
repeat\_region /note="AluJo repeat: matches 1. .302 of consensus"  
18800. .18956  
repeat\_region /note="MLTIA2 repeat: matches 1. .171 of consensus"  
19392. .19704  
repeat\_region /note="AluJo repeat: matches 1. .312 of consensus"  
19780. .20069  
repeat\_region /note="AluY repeat: matches 1. .290 of consensus"  
22285. .22456  
repeat\_region /note="MIR repeat: matches 23. .220 of consensus"  
22457. .22760  
repeat\_region /note="AluSg repeat: matches 1. .304 of consensus"  
22761. .22773  
repeat\_region /note="MIR repeat: matches 220. .230 of consensus"  
26734. .26961  
repeat\_region /note="L1MB2 repeat: matches 5944. .6170 of consensus"  
28074. .28174  
repeat\_region /note="MER44C repeat: matches 7. .103 of consensus"

repeat\_region 28175. .28840  
repeat\_region /note="MER72 repeat: matches 5. .668 of consensus"  
28841. .29431  
repeat\_region /note="MER44C repeat: matches 103. .721 of consensus"  
29821. .30105  
repeat\_region /note="AluX repeat: matches 1. .285 of consensus"  
30219. .30775  
repeat\_region /note="L2 repeat: matches 2150. .2749 of consensus"  
30924. .31207  
repeat\_region /note="AluX repeat: matches 3. .294 of consensus"  
32053. .32298  
repeat\_region /note="MIR repeat: matches 1. .257 of consensus"  
32498. .32740  
repeat\_region /note="L1MB3 repeat: matches 5943. .6182 of consensus"  
33867. .33920  
repeat\_region /note="27 copies 2 mer ca 90% conserved"  
33871. .33918  
repeat\_region /note="8 copies 6 mer cacaca 93% conserved"  
33878. .33921  
repeat\_region /note="11 copies 4 mer acac 97% conserved"  
34367. .34420  
repeat\_region /note="9 copies 6 mer ctcttc 75% conserved"  
34458. .34576  
repeat\_region /note="MIR repeat: matches 94. .219 of consensus"  
34914. .35006  
repeat\_region /note="MIR repeat: matches 150. .242 of consensus"  
35044. .35414  
repeat\_region /note="MLTIB repeat: matches 1. .386 of consensus"  
35440. .35480  
repeat\_region /note="L2 repeat: matches 2423. .2464 of consensus"  
35653. .36028  
repeat\_region /note="THEIB repeat: matches 1. .359 of consensus"  
36683. .37012  
repeat\_region /note="AluX repeat: matches 1. .308 of consensus"  
37526. .37818  
repeat\_region /note="AluJo repeat: matches 1. .290 of consensus"  
37957. .38028  
repeat\_region /note="36 copies 2 mer tg 86% conserved"  
37968. .38031  
repeat\_region /note="16 copies 4 mer gtgt 90% conserved"  
37980. .38027  
repeat\_region /note="8 copies 6 mer gtgtgt 100% conserved"  
38553. .39205  
repeat\_region /note="L1MB3 repeat: matches 5545. .6183 of consensus"  
39561. .39944  
repeat\_region /note="WSTA repeat: matches 1. .426 of consensus"  
40827. .41086  
repeat\_region /note="AluX repeat: matches 1. .259 of consensus"  
41793. .42089  
repeat\_region /note="AluSg repeat: matches 1. .297 of consensus"  
42330. .42393  
repeat\_region /note="L2 repeat: matches 2627. .2701 of consensus"  
42692. .42857  
repeat\_region /note="L1MC2 repeat: matches 5654. .5816 of consensus"  
42858. .43151  
repeat\_region /note="AluSg repeat: matches 1. .293 of consensus"  
43152. .43183  
repeat\_region /note="L1MC2 repeat: matches 5816. .5846 of consensus"  
43184. .43479  
repeat\_region

Query Match 37.9%; Score 107; DB 9; Length 96183;  
Best Local Similarity 100.0%; Pred. No. 4.2e-22;  
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 CAAAGAGCTCTGGGAAGTGTGCTGGATGCAGTTTGGCAACCACTCAACCTCGTGGAA 129  
|||||  
Db 44203 CAAAGAGCTCTGGGAAGTGTGCTGGATGCAGTTTGGCAACCACTCAACCTCGTGGAA 44262  
|||||  
QY 130 GTTGACTATTGGCCCTCGAGTTTCTCTGATCACAAAAGATCACGCT 176  
|||||  
Db 44263 GTTGACTATTGGCCCTCGAGTTTCTCTGATCACAAAAGATCACGCT 44309  
|||||

## RESULT 3

## AC022669/c

## LOCUS

AC022669 152053 bp DNA linear HTG 24-AUG-2002  
Homo sapiens chromosome 13 clone RP11-96B23 map 13, WORKING DRAFT  
SEQUENCE, 33 unordered pieces.

## ACCESSION

AC022669.4 GI:9966291

## VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## KEYWORDS

## SOURCE

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 152053)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 152053)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearliano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,  
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirelli,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

3 (bases 1 to 152053)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,  
Camata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,  
Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,  
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
Iliev,I., Johnson,R., Jones,C., Karatas,A., Horton,L., Hulme,W.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,  
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Ribback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,  
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S.,  
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Sep 4, 2000 this sequence version replaced gi:7229803.

All repeats were identified using RepeatMasker:

Smit,A.F.A. &amp; Green,P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4938

Center clone name: 96\_B\_23

----- Summary Statistics

Sequencing vector: M13; M77815; 93% of reads

Sequencing vector: Plasmid; n/a; %0-f% of reads

7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of

reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136451 bases at least Q40

Consensus quality: 143612 bases at least Q30

Consensus quality: 146365 bases at least Q20

Insert size: 158000; agarose-rp

Insert size: 148853; sum-of-ontigs

Quality coverage: 3.4 in Q20 bases; agarose-rp

Quality cov.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 33 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved. 784: contig of 784 bp in length

\* 1

\* 785 884: gap of 100 bp

\* 885 1956: contig of 1072 bp in length

\* 1957 2056: gap of 100 bp

\* 2057 3113: contig of 1057 bp in length

\* 3114 3213: gap of 100 bp

\* 3214 4817: contig of 1604 bp in length

\* 4818 4917: gap of 100 bp

\* 4918 6582: contig of 1665 bp in length

\* 6583 6682: gap of 100 bp

\* 6683 8059: contig of 1377 bp in length

\* 8060 8159: gap of 100 bp

\* 8160 22369: contig of 14210 bp in length

\* 22370 22469: gap of 100 bp

\* 22470 24974: contig of 2505 bp in length

\* 24975 25074: gap of 100 bp

\* 25075 27206: contig of 2132 bp in length

\* 27207 27306: gap of 100 bp

\* 27307 29420: contig of 2114 bp in length

\* 29421 29520: gap of 100 bp

\* 29521 32000: contig of 2480 bp in length

\* 32001 32100: gap of 100 bp

\* 32101 36330: contig of 4230 bp in length

\* 36331 36430: gap of 100 bp

\* 36431 39067: contig of 2637 bp in length

\* 39068 39167: gap of 100 bp

\* 39168 42564: contig of 3397 bp in length

\* 42565 42664: gap of 100 bp

\* 42665 46664: contig of 4000 bp in length

\* 46665 46764: gap of 100 bp

\* 46765 51047: contig of 4283 bp in length

\* 51048 51147: gap of 100 bp

\* 51148 55269: contig of 4122 bp in length

\* 55270 55369: gap of 100 bp

\* 55370 58831: contig of 3462 bp in length

\* 58832 58931: gap of 100 bp

\* 58932 64077: contig of 5146 bp in length

\* 64078 64177: gap of 100 bp

\* 64178 68808: contig of 4631 bp in length

\* 68809 68908: gap of 100 bp

\* 68909 74170: contig of 5262 bp in length

\* 74171 74270: gap of 100 bp

\* 74271 79296: contig of 5026 bp in length

\* 79297 79396: gap of 100 bp

\* 79397 84724: contig of 5328 bp in length

\* 84725 84824: gap of 100 bp

\* 84825 89713: contig of 4889 bp in length

\* 89714 89813: gap of 100 bp

\* 89814 94721: contig of 4908 bp in length

\* 94722 94821: gap of 100 bp

\* 94822 100196: contig of 5375 bp in length

\* 100197 100296: gap of 100 bp  
\* 100297 107015: contig of 6719 bp in length  
\* 107016 107115: gap of 100 bp  
\* 107116 114507: contig of 7392 bp in length  
\* 114508 114607: gap of 100 bp  
\* 114608 122671: contig of 8064 bp in length  
\* 122672 122771: gap of 100 bp  
\* 122772 128052: contig of 5281 bp in length  
\* 128053 128152: gap of 100 bp  
\* 128153 138583: contig of 10431 bp in length  
\* 138584 138683: gap of 100 bp  
\* 138684 150248: contig of 11565 bp in length  
\* 150249 150348: gap of 100 bp  
\* 150349 152053: contig of 1705 bp in length.

FEATURES

source  
1. .152053  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13"  
/clone\_lib="RPCI-11 Human Male BAC"  
misc\_feature  
1. .784  
/note="assembly\_fragment  
vector\_end:SP6  
885. .1956  
/note="assembly\_fragment"  
misc\_feature  
2057. .3113  
/note="assembly\_fragment"  
misc\_feature  
3214. .4817  
/note="assembly\_fragment"  
misc\_feature  
4918. .6582  
/note="assembly\_fragment"  
misc\_feature  
6683. .8059  
/note="assembly\_fragment"  
misc\_feature  
8160. .22369  
/note="assembly\_fragment"  
misc\_feature  
22470. .24974  
/note="assembly\_fragment"  
misc\_feature  
25075. .27206  
/note="assembly\_fragment"  
misc\_feature  
27307. .29420  
/note="assembly\_fragment"  
misc\_feature  
29521. .32000  
/note="assembly\_fragment"  
misc\_feature  
32101. .36330  
/note="assembly\_fragment"  
misc\_feature  
36431. .39067  
/note="assembly\_fragment"  
misc\_feature  
39168. .42564  
/note="assembly\_fragment"  
misc\_feature  
42665. .46664  
/note="assembly\_fragment"  
misc\_feature  
46765. .51047  
/note="assembly\_fragment"

Query Match 37.9%; Score 107; DB 2; Length 152053;  
Best Local Similarity 100.0%; Pred. No. 4.7e-22;  
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 CAAGAGCTCCTGGGAGGTGCTGTGATGCATTTGCAACCACTCAACCTCGTGGAA 129

|||||  
Db 31950 CAAGAGCTCCTGGGAGGTGCTGTGATGCATTTGCAACCACTCAACCTCGTGGAA 31891

QY 130 GGTGACTATTTGGCTCTGAGTTCTGTGATCACAAAAGATCAGGT 176

|||||  
Db 31890 GGTGACTATTTGGCTCTGAGTTCTGTGATCACAAAAGATCAGGT 31844

RESULT 4

BC021301

LOCUS

BC021301 2605 bp mRNA linear PRI 22-JAN-2002

DEFINITION

Homo sapiens, Similar to KIAA0793 gene product, clone MGC:29593  
IMAGE:5013180, mRNA, complete cds.

ACCESSION

BC021301

VERSION

BC021301.1 GI:18204274

KEYWORDS

MGC.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Strausberg, R.

REFERENCE

1 (bases 1 to 2605)

AUTHORS

Direct Submission

TITLE

Submitted (14-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK

NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT

Contact: MGC help desk

Email: cgabps@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Contact: nisc\_mgc@hgrl.nih.gov/  
Web site: http://www.nisc.nih.gov/

Contact: nisc\_mgc@hgrl.nih.gov

Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAL Plate: 39 Row: 1 Column: 18  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7662309.

FEATURES

Source

Location/Qualifiers  
1. .2605  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:29593 IMAGE:5013180"  
/tissue\_type="Brain, astrocytoma, grade IV"  
/clone\_lib="NIH\_MGC\_98"  
/lab\_host="DH10B"  
/note="Vector: pOTB7"  
126. .2069  
/codon\_start=1  
/product="Similar to KIAA0793 gene product"  
/protein\_id="AAH21301.1"  
/db\_xref="GI:16204275"  
/translation="MGEIEGTVRLQTACMRLGAQTPVGSTLEPGQTLLPRMQEHL  
HLRVKLLDNTMEIFDIEPKDQGLLTQVKNLNVCEDFGMEFQNTQSYTWLEPMP  
KPIIRRRKNVLRVLAFFPPDQLOEEVRYLFALQLKRLLERLCADTTPA  
ALLTSHLQSEIGDYDREHLKVNLYLPGQHCLEKILFEHQHVGQTPAESDPQ  
VLEIARKLEMYGIRHMASDREGTKIQLAVSHMGVLVFGCTTKINTFNWSKVRLSFK  
RRKRLKLHEVHGYPQDTLEFLLSRDECKNFWKICVETHTFFLLDDQPKPKAKAVF  
FRGSGFRYSGRGTQKQVDFYFKDSGMRIPYERRHSHKTHTSVRLTADLPKQISFPE  
GLRTPASPSANAFYSLSPGLTVPGLPEFKDSSSLTDPQVYKVPKPAARRSGAVA  
GGPDTPSAQPLPGPALPGPLSTKSPQSPSSRSKPLSLSPAFVPLPGPAGSGSPL  
LSPVLSDAGGAGMDCDEPRHKVPDAEAYFVKEILATERTYIKLDEVITVFRSAV  
KEDAPATLMLLEFNSIDPIEFHKGFLREVEQRLALWEGPSKAHTKGSHQRIGDILL  
RNRQLKVFQLHEGHVAGVTKE"

CDS

BASE COUNT 675 a 697 c 662 g 571 t  
ORIGIN  
Query Match 34.7%; Score 97.8; DB 9; Length 2605;  
Best Local Similarity 59.6%; Pred. No. 1.5e-19;  
Matches 165; Conservative 0; Mismatches 112; Indels 0; Gaps 0;







Assembly program: Phrap; version 0.990329  
 Consensus quality: 106401 bases at least Q40  
 Consensus quality: 113612 bases at least Q30  
 Consensus quality: 117819 bases at least Q20

\*\*\*\*\*  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 61 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 1127: contig of 1127 bp in length
* 1128 1227: gap of unknown length
* 1228 2235: contig of 1008 bp in length
* 2236 2335: gap of unknown length
* 2336 3792: contig of 1457 bp in length
* 3793 3892: gap of unknown length
* 3893 4977: contig of 1085 bp in length
* 4978 5077: gap of unknown length
* 5078 6106: contig of 1029 bp in length
* 6107 6206: gap of unknown length
* 6207 7617: contig of 1411 bp in length
* 7618 7717: gap of unknown length
* 7718 9305: contig of 1588 bp in length
* 9306 9405: gap of unknown length
* 9406 10632: contig of 1227 bp in length
* 10633 10732: gap of unknown length
* 10733 12158: contig of 1426 bp in length
* 12159 12258: gap of unknown length
* 12259 13281: contig of 1023 bp in length
* 13282 13381: gap of unknown length
* 13382 15064: contig of 1683 bp in length
* 15065 15164: gap of unknown length
* 15165 17118: contig of 2554 bp in length
* 17119 17718: gap of unknown length
* 17719 19363: contig of 1545 bp in length
* 19364 19463: gap of unknown length
* 19464 20526: contig of 1063 bp in length
* 20527 21864: contig of 1238 bp in length
* 21865 21964: gap of unknown length
* 21965 23838: contig of 1874 bp in length
* 23839 23938: gap of unknown length
* 23939 25034: contig of 1096 bp in length
* 25035 25134: gap of unknown length
* 25135 27370: contig of 2236 bp in length
* 27371 27470: gap of unknown length
* 27471 28692: contig of 1222 bp in length
* 28693 28792: gap of unknown length
* 28793 30073: contig of 1283 bp in length
* 30076 30176: gap of unknown length
* 30176 31471: contig of 1295 bp in length
* 31471 31570: gap of unknown length
* 31571 32895: contig of 1325 bp in length
* 32896 32995: gap of unknown length
* 32996 34801: contig of 1806 bp in length
* 34802 34901: gap of unknown length
* 34902 36662: contig of 1761 bp in length
* 36663 36762: gap of unknown length
* 36763 37997: contig of 1235 bp in length
* 37998 38097: gap of unknown length
* 38098 40561: contig of 2464 bp in length
* 40562 40661: gap of unknown length
* 40662 42524: contig of 1863 bp in length
* 42525 42624: gap of unknown length
* 42625 44247: contig of 1623 bp in length
* 44248 44347: gap of unknown length
* 44348 46510: contig of 2163 bp in length
* 46511 46610: gap of unknown length

```

```

* 46611 48669: contig of 2059 bp in length
* 48670 48769: gap of unknown length
* 48770 50218: contig of 1449 bp in length
* 50219 50318: gap of unknown length
* 50319 52454: contig of 2136 bp in length
* 52455 52554: gap of unknown length
* 52555 53956: contig of 1402 bp in length
* 53957 54056: gap of unknown length
* 54057 56174: contig of 2118 bp in length
* 56175 56274: gap of unknown length
* 56275 58482: contig of 2208 bp in length
* 58483 58582: gap of unknown length
* 58583 60511: contig of 1929 bp in length
* 60512 60611: gap of unknown length
* 60612 62512: contig of 1901 bp in length
* 62513 62612: gap of unknown length
* 62613 63352: contig of 2740 bp in length
* 63353 63452: gap of unknown length
* 63453 67770: contig of 2318 bp in length
* 67771 67870: gap of unknown length
* 67871 70795: contig of 2925 bp in length
* 70796 70895: gap of unknown length
* 70896 73397: contig of 2502 bp in length
* 73398 73497: gap of unknown length
* 73498 75528: contig of 2031 bp in length
* 75529 75628: gap of unknown length
* 75629 78574: contig of 2945 bp in length
* 78575 78673: gap of unknown length
* 78674 80883: contig of 2210 bp in length
* 80884 80983: gap of unknown length
* 80984 83541: contig of 2558 bp in length
* 83542 83641: gap of unknown length
* 83642 86220: contig of 2579 bp in length
* 86221 86320: gap of unknown length
* 86321 89057: contig of 2737 bp in length
* 89058 89157: gap of unknown length
* 89158 92362: contig of 3205 bp in length
* 92363 92462: gap of unknown length
* 92463 95949: contig of 3487 bp in length
* 95950 96049: gap of unknown length
* 96050 100396: contig of 4347 bp in length
* 100397 100496: gap of unknown length
* 100497 104325: contig of 3829 bp in length
* 104326 104425: gap of unknown length
* 104426 107324: contig of 2899 bp in length
* 107325 107424: gap of unknown length
* 107425 110542: contig of 3118 bp in length
* 110543 110642: gap of unknown length
* 110643 113543: contig of 2901 bp in length

```

Query Match 26.6%; Score 75; DB 2; Length 149417;  
 Best Local Similarity 81.3%; Pred No. 5.7e-12;

Matches 87; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 70 CAAGAGCTCTGGGAAGTGCTGCTGGATGAGTTTGCACACCTCAACCTCGTGAA 129

Db 35611 CAACGGCACCAGGAGGCTCTTTGATGCTGTGTGACCAACCTCAACCTTGTGGAG 35552

Qy 130 GGTGACTATTTGGCTCGAGTTTCTCTGATCATCAAAAAGATCACGGT 176

Db 35551 GGTGACTATTTGGCTCGAGTTTCTCTGATCATCAAAAAGATCGTGGT 35505

RESULT 8

AC122885/c

LOCUS AC122885

DEFINITION Mus musculus chromosome UNK clone RP23-90H14, WORKING DRAFT

ACCESSION AC122885

VERSION AC122885.2 GI:21539806

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE house mouse.

ORGANISM Mus musculus

221502 bp DNA linear HTG 23-JUN-2002

SEQUENCE, 8 unordered pieces.



VERHYARDQNNELLIGVMSGGILLIYNKVRMTFFLWLKIVKISFKCKQFFIOLRKEL  
HESRETLGFMNRYACKTLWACVEHHTFFRLRPLPPQKNFFAFTLGSKFYRC  
GRTEVOYQVKGKAKMDRVFARSSKGLHKLMDWEVVRNSLSDRLRTOQLSPRS  
PGTPNHRNSFTQEARVFRPSSGHLVDHVHMSPEDEVSQSPSTOANSIVLES  
SPQOETPEDQGPALPKQSKNSWQIHFESNQDLVTHNESFDVPSPKSTPNG  
GIPHNLVLKMKPDENRGFPNKGQYDKQMPVTVSRVAPGTPADLCVPRNEGDOV  
VLINGRDIAHHTDQVVLFIKASCEKHSGLVLLVRNAVIVVEKLESEPDFQYIP  
EKAPLDSVHDDHSLRESMTQALAEGLITGVLAQFDQLYRKFCMTMSKAPLONISK  
NRYRDISPYATRVLLKNGEDYINANYINNEIPSSYINOYIACQGPLHTCKDFWQM  
IWQSGSMVYMLTTOVERGRVKCHOYWPESSESSYGCYATCSHSEGNPAYIFRKWT  
LIQKRNESQLTQIQYTAMPDHGVDPDSSDFLDFVCHVDQVRAGEEPIIVHCSAGI  
GRGVLLTMTAMCLIECNQVPLDVRTRMDRAMMIQTPSQYRFVCEAILKYVEE  
GFVKPLTSSNK"

BASE COUNT 920 a 587 c 592 g 783 t  
ORIGIN  
Query Match 25.4%; Score 71.6; DB 10; Length 2882;  
Best Local Similarity 57.3%; Pred. No. 2.8e-11;  
Matches 150; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

QY 24 CAAMAATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTTCACAAAGAGCTCCTGG 83  
DB 108 CAACATCTTCTTCGGATACACTGTACAGCTTCAGAGTTACAAACATGATCAGG 167  
QY 84 GAAGTGTGCTGGATGAGTTGCAACACCTCAACCTCGTGAAGTGACTATTTTGG 143  
DB 168 GCAAGTCTGTGGATAGTCTTCAAGCATCTTGATTGACTGAGCGAGACTATTTTGG 227  
QY 144 CTTGAGTTTCT---GATCACAACAAATCAGGTGCGTGGATCCCTTAAACCCAT 200  
DB 228 TTTACAGTTGGCTGACGATTCACAGATACCAAGGTGGCTGATCCCAACAAACCAAT 287  
QY 201 TGTGAACAGATTGAAGCCCAACAGCGTTGTTTAAAGTTTGGTCAATTTCTTCC 260  
DB 288 AAGGAGCAGTAAGAGAGGATCACCITACAAATTGACITTAGAGTCAATTTCTTGT 347  
QY 261 GCCTGACCACACAACTCCAA 282  
DB 348 AAGTGACCCCAACAAAGTTACAA 369

RESULT 10 ALI36300 170916 bp DNA linear PRI 22-JUN-2001  
LOCUS Human DNA sequence from clone RP11-10G5 on chromosome 13 Contains a  
zinc finger gene similar to ZNF183, the start of the FARP1 (FERM,  
RHOGEF (ARHGEF) and plectstrin domain protein 1  
(chondrocyte-derived)) gene, ESTs, STSS, GSSs and CpG islands,  
complete sequence.  
ACCESSION ALI36300  
VERSION ALI36300.22 GI:11125392  
KEYWORDS HTG; FARP1; FERM; RHOGEF; zinc finger.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1. (bases 1 to 170916)  
AUTHORS Smith,M.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

COMMENT On Nov 8, 2000 this sequence version replaced gi:11061750.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information  
on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 13, constructed by the Sanger Centre Chromosome 13  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr13  
This sequence is the entire insert of clone RP11-10G5 The true left  
end of clone RP11-573N10 is at 109104 in this sequence. The true  
right end of clone RP11-72J7 is at 95284 in this sequence. This  
sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP11-10G5 is from the  
library RPCI-11.1 constructed by the group of Pieter de Jong. For  
further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6.

FEATURES

source  
1. .170916  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-10G5"  
/clone\_lib="RPCI-11.1"  
384. .873  
repeat\_region  
/note="MER31B repeat: matches 1. .481 of consensus"  
1374. .1407  
/note="17 copies 2 mer tt 85% conserved"  
5470. .5533  
/note="32 copies 2 mer tt 70% conserved"  
6352. .6714  
/note="match: GSS: Em:AQ698072"  
8498. .8665  
/note="TIGGER1 repeat: matches 2080. .2238 of consensus"  
8670. .8888  
/note="MER75B repeat: matches 18. .239 of consensus"  
13335. .13686  
/note="176 copies 2 mer tg 69% conserved"  
13996. .14238  
/note="L1PB3 repeat: matches 5894. .6150 of consensus"  
15610. .15639  
/note="15 copies 2 mer ta 86% conserved"  
complement(17881. .18076)  
/note="match: GSS: Em:AQ776608"  
18339. .18804  
/note="LTR7 repeat: matches 1. .448 of consensus"  
19891. .20105  
/note="MER57-internal repeat: matches 7171. .7387 of  
consensus"  
20131  
/note="Tandem repeat. Forced join. Gap size estimated to  
be approximately 260bp by restriction digest data."  
29106. .29251  
/note="LTR41 repeat: matches 2. .160 of consensus"  
31944. .32589  
/note="LTR12 repeat: matches 1. .671 of consensus"  
32590. .33323  
/note="P785 repeat: matches 743. .1478 of consensus"  
38726. .38851  
/note="63 copies 2 mer ca 88% conserved"  
41206. .41622  
/note="match: GSS: Em:AQ184690"  
42604. .43109  
/note="match: GSS: Em:AQ573113"  
43362. .43429  
/note="24 copies 2 mer ta 93% conserved"  
complement(46659. .47093)  
/note="match: GSS: Em:AQ805998"  
46717. .46754  
/note="19 copies 2 mer gt 100% conserved"  
53170. .53274  
repeat\_region

repeat\_region /note="MER34 repeat: matches 8. .115 of consensus"  
53278. .53412  
repeat\_region /note="MER74B repeat: matches 50. .180 of consensus"  
55996. .56438  
repeat\_region /note="MER83 repeat: matches 1. .448 of consensus"  
59536. .60314  
repeat\_region /note="MER21B repeat: matches 63. .794 of consensus"  
60344. .60392  
repeat\_region /note="MER21B repeat: matches 741. .790 of consensus"  
60825. .61336  
repeat\_region /note="MER74B repeat: matches 61. .621 of consensus"  
61616. .61665  
repeat\_region /note="MER61B repeat: matches 377. .425 of consensus"  
61667. .62339  
repeat\_region /note="LTR9 repeat: matches 1. .625 of consensus"  
62342. .62683  
repeat\_region /note="MER61C repeat: matches 1. .384 of consensus"  
63996. .64135  
repeat\_region /note="MLT2B repeat: matches 1. .149 of consensus"  
64486. .64768  
repeat\_region /note="MLT2E repeat: matches 149. .395 of consensus"  
65150. .65369  
repeat\_region /note="L1MA5A repeat: matches 6053. .6291 of consensus"  
65371. .66037  
repeat\_region /note="LTR12 repeat: matches 1. .647 of consensus"  
66038. .66088  
repeat\_region /note="L1MA5A repeat: matches 6037. .6091 of consensus"  
66076. .66299  
repeat\_region /note="L1MA5A repeat: matches 5827. .6047 of consensus"  
67351. .67427  
repeat\_region /note="MLT2E repeat: matches 173. .250 of consensus"  
67428. .67487  
repeat\_region /note="30 copies 2 mer ag 88% conserved"  
67492. .67660  
repeat\_region /note="MLT1A2 repeat: matches 5. .174 of consensus"  
68268. .68810  
repeat\_region /note="MER68A repeat: matches 1. .561 of consensus"  
complement(73998. .74451)  
repeat\_region /note="match: GSS: Em:AQ695161"  
74143. .74773  
repeat\_region /note="LTR39 repeat: matches 172. .794 of consensus"  
75063. .75233  
repeat\_region /note="LTR39 repeat: matches 1. .172 of consensus"  
79181. .79291  
repeat\_region /note="MLT1A2 repeat: matches 251. .374 of consensus"  
79699. .80013  
repeat\_region /note="MLT1A2 repeat: matches 1. .251 of consensus"  
80095. .80463  
repeat\_region /note="THE1C repeat: matches 1. .369 of consensus"  
83250. .83460  
repeat\_region /note="rigger3b repeat: matches 1. .210 of consensus"  
83757. .84298  
repeat\_region /note="rigger3b repeat: matches 210. .750 of consensus"  
84581. .84607  
repeat\_region /note="rigger3b repeat: matches 750. .775 of consensus"  
84608. .84963  
repeat\_region /note="THE1C repeat: matches 4. .371 of consensus"  
84964. .85004  
repeat\_region /note="rigger3b repeat: matches 775. .825 of consensus"  
85677. .86088  
repeat\_region /note="rigger3b repeat: matches 823. .1231 of consensus"  
complement(89463. .90009)  
repeat\_region /note="match: GSS: Em:AQ013482"  
complement(89490. .90015)  
repeat\_region /note="match: GSS: Em:AQ383425"  
complement(89526. .90008)  
repeat\_region /note="match: GSS: Em:B82154"  
complement(93764. .94093)  
repeat\_region /note="match: GSS: Em:AQ594661"  
98631. .100473  
misc\_feature /note="CpG island"  
99666. .99802

misc\_feature /note="match: GSS: Em:AQ938343"  
99694. .99750  
repeat\_region /note="single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."  
100889. .100712  
misc\_feature /note="12 copies 2 mer tg 100% conserved"  
101250. .101669  
misc\_feature /note="match: GSS: Em:AQ008502"  
101284. .101479  
repeat\_region /note="match: GSS: Em:AQ000309"  
105297. .105359  
misc\_feature /note="rLMEC repeat: matches 1638. .1710 of consensus"  
108760. .109211  
misc\_feature /note="match: GSS: Em:B66665"  
108771. .109149  
misc\_feature /note="match: GSS: Em:AQ076004"  
108818. .109537  
repeat\_region /note="match: GSS: Em:AQ307265"  
109801. .110016  
misc\_feature /note="108 copies 2 mer at 58% conserved"  
117543. .118068  
repeat\_region /note="match: GSS: Em:AQ678189"  
118206. .118247  
Query Match . 25.1%; Score 70.8; DB 9; Length 170916;  
Best Local Similarity 97.3%; Pred. No. 1.2e-10;  
Matches 72; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 COTTCAGGAAACTCGTTCATCAAAATCCAGATGCTGGATGACACCCAGAGGCATT 60  
|||||  
Db 169831 COTTCAGGAAACTCGTTCATCAAAATCCAGATGCTGGATGACACCCAGAGGCATT 169890  
QY 61 GAAGTTCACCAAG 74  
|||||  
Db 169891 GAAGTTCACGTAAG 169904  
RESULT 11  
AB032366  
LOCUS AB032366 3451 bp mRNA linear ROD 23-MAY-2000  
DEFINITION Mus musculus Ehm2 mRNA, complete cds.  
ACCESSION AB032366  
VERSION AB032366.1 GI:8051681  
KEYWORDS EHM2.  
SOURCE Mus musculus cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (sites)  
AUTHORS Hashimoto,Y., Shindo-Okada,N., Tani,M., Takeuchi,K., Toma,H. and Yokota,J.  
TITLE Identification of genes differentially expressed in association with metastatic potential of K-1735 murine melanoma by messenger RNA differential display  
JOURNAL Cancer Res. 56 (22), 5266-5271 (1996)  
MEDLINE 97069887  
REFERENCE 2 (sites)  
AUTHORS Shimizu,K., Nagamachi,Y., Tani,M., Kimura,K., Shiroidshi,T., Wakana,S. and Yokota,J.  
TITLE Molecular cloning of a novel NF2/ERM/4.1 superfamily gene, ehm2, that is expressed in high-metastatic K1735 murine melanoma cells  
JOURNAL Genomics 65 (2), 113-120 (2000)  
MEDLINE 20247250  
REFERENCE 3 (bases 1 to 3451)  
AUTHORS Yokota,J., Shimizu,K. and Nagamachi,Y.  
TITLE Direct Submission  
JOURNAL Submitted (04-SEP-1999) Jun Yokota, National Cancer Center Research Institute, Biology Division; Tsukiji 5-chome 1-1, Chuo-ku, Tokyo 104-0045, Japan (E-mail:jyokota@ncc.ncc.go.jp,  
Tel:81-3-3547-5272, Fax:81-3-3542-0807)  
FEATURES  
1. .3451  
source

	BASE COUNT
gene	764 a 958 c 967 g 762 t
CDS	ORIGIN

Query Match	25.08;	Score	70.4;	DB	10;	Length	3451;
Best Local Similarity	55.2%;	Pred.	No.	6.9e-11;			
Matches	137;	Conservative	0;	Mismatches	111;	Indels	0;
Qy	35	TGCTGGATGACACCCAGGAGGCATTTGAAGTTCACAAAGAGCTCTCGGAAGGTGCTGC	94				
Db	770	TGCTCGACGGAACCGAAGTGAGCGCTGGACCTGCGCGAAACATGCCAAAGGCCAGGATTGT	829				
Qy	95	TGGATGCAGTTTGCAACACCACCTCAACCTCGTGGGAAGTGACTATTTTGGCTCGAGTTTC	154				
Db	830	TCGACCAGATTGTGTACCACCTTGGAACCTGGTGGGAACAGAGTATTTTGGTCTGCAGTTCC	889				
Qy	155	CTGATCAAAAAGATCACGGTGTGGCTGCGATCTCTAAAACCCATTGTGAAACAGATTA	214				
Db	890	TCGACTCTGCCAGGTACGCACCTGGCTGGATCAGCGGAACCCCATTAAGAAACAGATGA	949				
Qy	215	GAAGGCCAAAGCAGTGTGTTTAAAGTTTGTGGTGAATTTCTTTCCGCCTGACCACACAC	274				
Db	950	AAGTGGGACCTGCCTATGCCTTGGACITTCGAGTCANAATACTATTCTTCGAACCAACA	1009				
Qy	275	AACTCCAA	282				
Db	1010	ACCTTCGA	1017				

RESULT 12	
BC010674	
LOCUS	3619 bp mRNA linear PRI 12-JUL-2001
DEFINITION	Homo sapiens, protein tyrosine phosphatase, non-receptor type 4 (megakaryocyte), clone MGC:9204 IMAGE:3853914, mRNA, complete cds.
ACCESSION	BC010674
VERSION	BC010674.1 GI:14715026
KEYWORDS	MGC.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 3619)
TITLE	Strausberg, R.
JOURNAL	Direct Submission
	Submitted (10-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: villalona@bcm.tmc.edu  
 Villalón, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., García,  
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
 Muzny, D.W., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 13 Row: b Column: 15  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA qi: 190747.

	FEATURES	SOURCE
1.	100% Cotton	USA
2.	Machine Washable	USA
3.	Soft Touch	USA
4.	Durable Construction	USA
5.	Available in Multiple Colors	USA
6.	Easy Care	USA
7.	Comfortable Fit	USA
8.	Long-Lasting Quality	USA
9.	Stylish Design	USA
10.	Wide Range of Sizes	USA

CDS

GFVKPLTTSNK"			
BASE COUNT	1116 a	762 c	789 g
ORIGIN			
			952 t

Query Match	24.3%	Score 68.4;	DB 9;	Length 3619;
Best Local Similarity	56.5%;	Pred. No. 3e-10;		
Matches 148;	Conservative 0;	Mismatches 111;	Indels 3;	Gaps 1;
QY	24	CAAAATCCAGATGCTGCATGACACCCAGGAGGCATTTGAAGTTCACAAAGAGCTCCTGG	83	
Db	520	CACATCCTTCTCTGGATACACTGTACAGCTCTTCAAAAGTCATTAAGCATGATCAGGG	579	
QY	84	GAAGTGCTGTGGATGCAGTTTGCAACACCACCTCAACCTCGTGGAGGTGACTATTTTGG	143	
Db	580	GCAAGTCTGTGGATGCTGCTCTCAAGCATCTAGATTTGACTGAGCAGGACTATTTTGG	639	
QY	144	CTCGAGTTTCTCGATCA---CAAAAAGATACGGTGTGGTGGATCTCTAAACCCAT	200	
Db	640	TTTACAGTCTGGCTGATGATTTCCACAGATAACCCAAAGTGGCTGGATCTCAACAAACCAAT	699	
QY	201	TGTGAACACAGATTAGAAGGCCAAAGCAGCTGTGTGTTAAAGTTTCTGCTGAAATTCCTTCC	260	
Db	700	AAGGAACGACGTAAGAGAGGATCTCCTTACAGTTTGAACTTTAGAGTCAAATTTTTTGT	759	
QY	261	GCCTGACCACACACACTCCAA	282	

Db 760 AAGTGACCCCAACAGTTTACA 781

RESULT 13  
HUMPTYPH  
LOCUS Human protein-tyrosine phosphatase mRNA, complete cds.  
DEFINITION 3643 bp mRNA linear PRI 27-APR-1993  
ACCESSION M88941  
VERSION M88941.1 GI:190747  
KEYWORDS protein-tyrosine phosphatase.  
SOURCE Homo sapiens (library: cDNA, Meg-01, HUVEC) cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Gu,M.X., York,J.D., Marshawsky,I. and Majerus,P.W.  
IDENTIFICATION, cloning, and expression of a cytosolic  
megakaryocyte protein-tyrosine-phosphatase with sequence homology  
to cytoskeletal protein.4.1  
Proc. Natl. Acad. Sci. U.S.A. 88 (13), 5867-5871 (1991)  
JOURNAL 91288564  
MEDLINE PUBMED 1648233  
FEATURES  
source Location/Qualifiers  
1..3643  
    /organism="Homo sapiens"  
    /db\_xref="taxon:9606"  
    /cell\_line="Meg-01, human platelets"  
    /tissue\_lib="cDNA, Meg-01, HUVEC"  
    <1..3643  
        /product="protein-tyrosine phosphatase"  
1..771  
    /note="contains short ORF's, GC rich stem loop structure  
with 75% homology to cABL protooncogene 5'UTR; putative"  
632..673  
    /note="GC rich stem loop structure"  
772..3552  
    /codon\_start=1  
    /product="protein-tyrosine phosphatase"  
    /protein\_id="AAA36530.1"  
    /db\_xref="GI:190748"  
    /translations="MISRRLPAGTVNVVASELARDRQHTVEVCNLLILLDNTVQAQPS  
VNKHDOGVLLDVVFVKHLDLTEQDFGLQLADSTPNPLWDNPKPIRKOLKKGSPYS  
LNFVRFFVSDPNKLGEYTRYQFYLIQKDILTRCPNSTAALFAASFVAOSELGID  
YDSLENLSGLVSYSEIPNQDFEXEIAKHOOHIGLSPAEAFNYLNRTARTLELYG  
VEFHARDQNIMIGVMNSGGIILYKNRYMTNPWLKIVISFKCKQFFIQLRKE  
HESRETLGFMVNYRNACKNLKACVEHHTFFRLDRPLPKDKMFKFAHYFTLGSKFRYC  
GRTEVOVSQYGGKAKNDRVFARPSKPLARKLMDEWVVSRSNISDDRLETOSLPERS  
PCGTGNHRNSTFTQEGRTPAPSVGHVLVHMVITSPSEFVNORSSTQANSIVLES  
SPSGCTPGDGKPALPPKQSKNSNQIHYSHQDDLESINETFDIPISPKEPTPNG  
GIPHNDLVLRMPKEDRGFNFGKYGDQMPVIVSRVAPGTPADLCVPRNLNEGQVIP  
VLINGDLIAETHDQVLFETKASCRHSGLMLLVPRNAVYDVVEEKLENEPOFIPI  
EKAPLSDVHODDSLRESMIQLAEGLTIGVLTFQDLVRKKPGMTMSACKLPQNIKS  
NRYRDISPYDATRIKLNEDYINANYINWEIPSSIIINQYIACQGPLPHCTCDFMQM  
TWQGSWMVMLTQTVERGRVKCHQWPETGSSSYGCTQVICHSEGNTAYIFRRMT  
LFWQENESRPLTQIQI IAWPDHGVDSSDDLDFVCHVRNKRAKKEEPVVVHCSAIG  
GRTGVLDTMETAMCLIECNPVYLDIRVTMRDQRMMIQTSPSQYRFVCEAILKVVEE  
GFVKPLTTSTNK"

BASE COUNT 1038 a 815 c 863 g 927 t  
ORIGIN

Query Match 24.3%; Score 68.4; DB 9; Length 3643;  
Best Local Similarity 56.5%; Pred. No. 3e-10;  
Matches 146; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

Qy 24 CAAATCCAGATCGTGATGACACCCAGGAGCATTTGAAGTCCACAAAGACTCTCTGG 83  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 864 CAACATCCTCTTCGTGATAACACTGTACAGCTTTTCAAAGTCAATAAAGATGATCAGG 923  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 84 GAAGTGCTGCTGCATGCAGTTTGCACACCACCTCAACCTCGTGGAGGTGACTATTGG 143  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 924 GCAAGTCTTGTGGATGCTGCTTCACAGCATCTAGATTTGACTGAGCAGGACTATTGG 983  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 144 CCTCGATTTTCTTGATCA---CAAAAAAGATCACGGTGTGGCTGGATCTCTCTAAACCCAT 200

Db	984	TTTACAGTTGGCTGATGATTCACAGATAACCCAGGTGGCTGGATCCAAACAAACCAAT	1043
Qy	201	TGTGAACACAGATTAGAAGGCCAAAGCAGTGTCTTAAAGTTTGTGGTGAATTCCTTTCC	260
Db	1044	AAGNAGCAGCTAAGAGAGAGATCTCCTTACAGATTGAACCTTGTAGAGTCAATATTTTGT	1103
Qy	261	GCCTGACCACACACAACTCCAA	282
Db	1104	AAGTGACCCCAACAAAGTTACAA	1125
RESULT 14			
AC022669			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
RESULT 14			
AC022669			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 4, 2000 this sequence version replaced gi:7229803.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L4938  
Center clone name: 96\_B\_23  
----- Summary Statistics  
Sequencing vector: M13: M7815; 93% of reads  
Sequencing vector: Plasmid; n/a; %0.f% of reads  
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 136451 bases at least Q40  
Consensus quality: 143612 bases at least Q30  
Consensus quality: 146365 bases at least Q20  
Insert size: 158000; agarose-fp  
Insert size: 148853; sum-of-contigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality cov.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 784: contig of 784 bp in length  
\* 785 884: gap of 100 bp  
\* 885 1956: contig of 1072 bp in length  
\* 1957 2056: gap of 100 bp  
\* 2057 3113: contig of 1057 bp in length  
\* 3114 3213: gap of 100 bp  
\* 3214 4817: contig of 1604 bp in length  
\* 4818 4917: gap of 100 bp  
\* 4918 6582: contig of 1665 bp in length  
\* 6583 8059: contig of 1377 bp in length  
\* 8060 8159: gap of 100 bp  
\* 8160 22369: contig of 14210 bp in length  
\* 22370 22469: gap of 100 bp  
\* 22470 24974: contig of 2505 bp in length  
\* 24975 25074: gap of 100 bp  
\* 25075 27206: contig of 2132 bp in length  
\* 27207 27306: gap of 100 bp  
\* 27307 29420: contig of 2114 bp in length  
\* 29421 29520: gap of 100 bp  
\* 29521 32000: contig of 2480 bp in length  
\* 32001 32100: gap of 100 bp  
\* 32101 36330: contig of 4230 bp in length  
\* 36331 36430: gap of 100 bp  
\* 36431 39067: contig of 2637 bp in length  
\* 39068 39167: gap of 100 bp  
\* 39168 42564: contig of 3397 bp in length  
\* 42565 42664: gap of 100 bp  
\* 42665 46664: contig of 4000 bp in length  
\* 46665 46764: gap of 100 bp  
\* 46765 51047: contig of 4283 bp in length  
\* 51048 51147: gap of 100 bp  
\* 51148 55269: contig of 4122 bp in length  
\* 55270 55369: gap of 100 bp  
\* 55370 58831: contig of 3462 bp in length  
\* 58832 58931: gap of 100 bp  
\* 58932 64077: contig of 5146 bp in length  
\* 64078 64177: gap of 100 bp  
\* 64178 68808: contig of 4631 bp in length  
\* . . .

\* 68809 68908: gap of 100 bp  
\* 68909 74170: contig of 5262 bp in length  
\* 74171 74270: gap of 100 bp  
\* 74271 79296: contig of 5026 bp in length  
\* 79297 79396: gap of 100 bp  
\* 79397 84724: contig of 5328 bp in length  
\* 84725 84824: gap of 100 bp  
\* 84825 89713: contig of 4889 bp in length  
\* 89714 89813: gap of 100 bp  
\* 89814 94721: contig of 4908 bp in length  
\* 94722 94821: gap of 100 bp  
\* 94822 100196: contig of 5375 bp in length  
\* 100197 100296: gap of 100 bp  
\* 100297 107015: contig of 6719 bp in length  
\* 107016 107115: gap of 100 bp  
\* 107116 114507: contig of 7392 bp in length  
\* 114508 114607: gap of 100 bp  
\* 114608 122671: contig of 8064 bp in length  
\* 122672 122771: gap of 100 bp  
\* 122772 128052: contig of 5281 bp in length  
\* 128053 128152: gap of 100 bp  
\* 128153 138583: contig of 10431 bp in length  
\* 138584 138683: gap of 100 bp  
\* 138684 150248: contig of 11565 bp in length  
\* 150249 150348: gap of 100 bp  
\* 150349 152053: contig of 1705 bp in length.  
\* Location/Qualifiers  
1..152053  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13"  
/clone="RP11-96B23"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..784  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
885..1956  
/note="assembly\_fragment"  
2057..3113  
/note="assembly\_fragment"  
3214..4817  
/note="assembly\_fragment"  
4918..6582  
/note="assembly\_fragment"  
6683..8059  
/note="assembly\_fragment"  
8160..22369  
/note="assembly\_fragment"  
22470..24974  
/note="assembly\_fragment"  
25075..27206  
/note="assembly\_fragment"  
27307..29420  
/note="assembly\_fragment"  
29521..32000  
/note="assembly\_fragment"  
32101..36330  
/note="assembly\_fragment"  
36431..39067  
/note="assembly\_fragment"  
39168..42564  
/note="assembly\_fragment"  
42665..46664  
/note="assembly\_fragment"  
46765..51047  
/note="assembly\_fragment"

Query Match 23.2%; Score 65.4; DB 2; Length 152053;  
Best local Similarity 98.5%; Pred. No. 6e-09; Indels 0;  
Matches 66; Conservative 0; Mismatches 1; Gaps 0;





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 14:48:04 ; Search time 44.5773 Seconds  
(without alignments)  
14246.337 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_151\_432

Perfect score: 282

Sequence: 1 ccttcaggaaactcgtgtc.....ctgaccacacacaaactccaa 282

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	282	100.0	3187	23 AAS64760	DNA encoding novel
2	282	100.0	3442	20 AAX79183	Human chondrocyte-
3	282	100.0	4687	21 AAA08582	Human cytoskeleton
c 4	138	48.9	241	AAA42150	Human secreted exp
5	68.4	24.3	3643	24 ABK84074	Human cDNA differe
6	65.2	23.1	659	22 AAH33845	Human colon cancer
7	65.2	23.1	1234	23 AAS73863	DNA encoding novel
8	65.2	23.1	2872	22 AAS08887	Human protein tyro
9	65.2	23.1	3313	23 ABN93400	Human gene GS91552

10	61.4	21.8	14917	24 AAD32338	Human lung specifi
11	60.6	21.5	3106	23 ABN93399	Human gene GS91552
12	57.8	20.5	621	24 ABQ72602	Human MDDT encodin
13	57.8	20.5	2846	23 ABL60869	Human cytoskeleton
14	57.8	20.5	4922	24 ABV30214	Human prostate exp
15	56.8	20.1	6263	24 ABK70019	Human cDNA for the
16	56.6	20.1	3031	23 ABV22627	Human prostate exp
17	56.6	20.1	3031	23 ABV28450	Human prostate exp
18	54.6	19.4	3984	18 AAT58627	Protein tyrosine p
19	54.6	19.4	3984	20 AAX78463	Human PTPH1 CDNA.
20	54.6	19.4	3984	24 AAT72918	PTPH1 CDNA. Homo
21	54	19.1	550	22 AAH08244	Human cDNA clone (
22	54	19.1	2067	24 ABLN85314	Human cytoskeleton
23	54	19.1	3166	22 AAH17787	Human cDNA sequenc
24	54	19.1	4356	23 ABL29969	Drosophila melanog
25	52.8	18.7	556	22 AAH99189	Human protein enco
26	47.6	16.9	3620	21 AAA08581	Human cytoskeleton
27	46.6	16.5	2337	24 ABQ73378	Human red blood ce
28	45.8	16.2	1626	22 AAI59140	Human polynucleoti
29	45.8	16.2	1685	24 ABL50340	Human cancer cell
30	45.8	16.2	2556	22 AAI60926	Human polynucleoti
31	45	16.0	1220	23 ABL02335	Drosophila melanog
32	44.6	15.8	347	21 AAC02360	Human secreted pro
c 33	42.6	15.1	8105	23 ABL29968	Drosophila melanog
34	41.8	14.8	427	21 AAC01157	Human secreted pro
35	39.2	13.9	650	22 AAS31134	Human diagnostic a
36	39.2	13.9	2066	24 ABL60867	Human cytoskeleton
c 37	38.6	13.7	414	24 ABL66151	Lung cancer relate
38	37.2	13.2	336	21 AAC01319	Human secreted pro
39	37.2	13.2	512	24 ABK70156	Human lung cancer
40	37.2	13.2	562	21 AAA69837	Human ovarian carc
41	37.2	13.2	562	24 ABLN7231	Ovarian carcinoma
42	37.2	13.2	2930	24 ABQ88181	Human osteoblast d
43	37.2	13.2	2930	24 ABK70285	Human lung cancer
44	37.2	13.2	3044	24 ABQ88180	Human osteoblast d
45	37.2	13.2	3044	24 ABK84552	Human cDNA differe

## ALIGNMENTS

## RESULT 1

AAS64760

ID AAS64760 standard; cDNA; 3187 BP.

AC AAS64760;

XX AAS64760;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #564.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG00573.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PS Claim 1; SEQ ID No 564; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 3187 BP; 810 A; 874 C; 876 G; 527 T; 0 other;

Query Match 100.0%; Score 282; DB 23; Length 3187;

Best Local Similarity 100.0%; Pred. No. 3.9e-83;

Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60  
DB 151 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 210  
QY 61 GAAGTTCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGCAATTCGCAACACCTCAAC 120  
DB 211 GAAGTTCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGCAATTCGCAACACCTCAAC 270  
QY 121 CTCGTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCAGAAAAGATCACGGTGTGG 180  
DB 271 CTCGTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCAGAAAAGATCACGGTGTGG 330  
QY 181 CTGGATCTCTAAACCCATTTGGAACAGATTAGAGGCCAAGACACGTTGTTGTTAAG 240  
DB 331 CTGGATCTCTAAACCCATTTGGAACAGATTAGAGGCCAAGACACGTTGTTGTTAAG 390  
QY 241 TTTGTGGTGAATTTCTTCCGCTGACACACACAACTCCAA 282  
DB 391 TTTGTGGTGAATTTCTTCCGCTGACACACACAACTCCAA 432

RESULT 2

AA79183

ID AA79183 standard; DNA; 3442 BP.

XX

AC AA79183;

XX

DT 17-AUG-1999 (first entry)

XX

DE Human chondrocyte-derived gene CDEP.

XX

KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;  
KW Dbl homology domain; pleckstrin homology domain; rheumatoid 'arthritis;  
KW drug; ss.

XX

OS Homo sapiens.

XX

XX WO928458-A1.

XX

PD 10-JUN-1999.  
XX  
PF 27-NOV-1998; 98WO-JP05348.  
XX  
PR 27-NOV-1997; 97JP-0342060.  
XX  
PA (CHUS ) CHUGAI SEIYAKU KK.  
XX  
PI Kato Y, Kawamoto T, Koyano Y;  
XX  
XX WPI; 1999-371117/31.  
DR P-PSDB; AAY07482.  
XX

PT Protein CDEP expressed in differentiated chondrocytes, and gene  
PT encoding it  
XX

PS Claim 5; Fig 1; 59pp; Japanese.

XX This sequence represents the coding region for a protein (CDEP) expressed  
CC in differentiated human foetal chondrocytes, which contains an ezrin-like  
CC domain, a Dbl homology (DH) domain and a pleckstrin homology (PH) domain.  
CC The nucleic acid or protein can be used in the investigation and  
CC treatment of cancers and arthritic diseases (including chronic rheumatoid  
CC arthritis), or for screening of candidate anticancer drugs.

XX  
SQ Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;

Query Match 100.0%; Score 282; DB 20; Length 3442;

Best Local Similarity 100.0%; Pred. No. 4.1e-83;

Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60  
DB 151 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 210  
QY 61 GAAGTTCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGCAATTCGCAACACCTCAAC 120  
DB 211 GAAGTTCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGCAATTCGCAACACCTCAAC 270  
QY 121 CTCGTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCAGAAAAGATCACGGTGTGG 180  
DB 271 CTCGTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCAGAAAAGATCACGGTGTGG 330  
QY 181 CTGGATCTCTAAACCCATTTGGAACAGATTAGAGGCCAAGACACGTTGTTGTTAAG 240  
DB 331 CTGGATCTCTAAACCCATTTGGAACAGATTAGAGGCCAAGACACGTTGTTGTTAAG 390  
QY 241 TTTGTGGTGAATTTCTTCCGCTGACACACAACTCCAA 282  
DB 391 TTTGTGGTGAATTTCTTCCGCTGACACACAACTCCAA 432

RESULT 3

AAA08582

ID AAA08582 standard; DNA; 4687 BP.

XX

AC AAA08582;

XX

DT 19-JUL-2000 (first entry)

XX

DE Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.

XX

KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;

KW autoimmunity; inflammatory; vesicle trafficking; neurological;

KW cardiovascular; cell motility; reproductive; muscle disorder; ss.

XX

OS Homo sapiens.

XX

FH Key

FT CDS

FT

FT

XX

Location/Qualifiers  
90..327  
/\*tag= a  
/product= CYSKP-2

PN WO200017355-A2.  
XX  
PD 30-MAR-2000.  
XX  
PF 17-SEP-1999; 99WO-US21565.  
XX  
XX 18-SEP-1998; 98US-0172226.  
PR 27-APR-1999; 99US-0131321.  
XX  
XX (INCY-) INCYTE PHARM INC.  
PI Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;  
PI Guegler KJ, Patterson C, Azimzal Y, Baughn MR;  
XX  
DR WPI: 2000-283582/24.  
XX P-PSDB: AAY91947.  
XX  
PT Human cytoskeleton associated proteins, used to treat cell  
PT proliferative, autoimmune/inflammatory, vesicle trafficking,  
PT neurological, cell motility, reproductive and muscle disorders  
XX  
PS Claim 9; Page 101-102; 113pp; English.  
XX  
XX AAA08581-96 encode human cytoskeleton associated proteins 1 to 16  
CC (CysRP-1 to CysRP-16) respectively. The sequences can be used to treat  
CC and diagnose cancer and cell proliferative, autoimmune/inflammatory,  
CC vesicle trafficking, neurological, cardiovascular, cell motility,  
CC reproductive and muscle disorders. Pharmaceutical compositions  
CC containing CysRP-2 to CysRP-16 can be used to treat or prevent disorders  
CC associated with decreased expression or activity of CysRP (claimed), for  
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,  
CC cancers, autoimmune/anti-inflammatory disorders such as allergies, anemia,  
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,  
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma  
CC and trauma. CysRP antagonists can be used to treat or prevent a  
CC disorder associated with increased expression or activity of CysRP  
CC (claimed).  
XX  
SQ Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;  
Query Match 100.0%; Score 282; DB 21; Length 4687;  
Best Local Similarity 100.0%; Pred. No. 4.7e-83;  
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTTCAGAAACTCGTGTCCATCAAAATCCAGATGCTGATGACACCCAGAGGCATT 60  
Db 192 CCTTCAGAAACTCGTGTCCATCAAAATCCAGATGCTGATGACACCCAGAGGCATT 251  
QY 61 GAAGTTCCCAAGAGCTCTCGGAAGGTGCTGCTGATGCGATTTGCAACCTCAAC 120  
Db 252 GAAGTTCCCAAGAGCTCTCGGAAGGTGCTGCTGATGCGATTTGCAACCTCAAC 311  
QY 121 CTGCTGGAAGTGACTATTTTGGCTGCTGAGTTTCTGATCACAAGATCAGCGTGG 180  
Db 312 CTGCTGGAAGTGACTATTTTGGCTGCTGAGTTTCTGATCACAAGATCAGCGTGG 371  
QY 181 CTGATCTCTAAACCCATTGTGAACAGATTAGAGGCCAAGCAGTGTGTGTTAAG 240  
Db 372 CTGATCTCTAAACCCATTGTGAACAGATTAGAGGCCAAGCAGTGTGTGTTAAG 431  
QY 241 TTTGTGTGAAATCTTTCCGCTGACCCACACAACTCAA 282  
Db 432 TTTGTGTGAAATCTTTCCGCTGACCCACACAACTCAA 473

RESULT 4  
AAA42150/C  
ID AAA42150 standard; cDNA; 241 BP.  
XX  
XX AAA42150;  
XX  
XX 21-AUG-2000 (first entry)  
DT  
XX

DE Human secreted expressed sequence tag SEQ ID NO:890.  
XX  
XX Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;  
KW expressed sequence tag; EST; probe; chemotactic; proliferative;  
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;  
KW antitumor; osteopathic; neuroprotective; nootropic; antiproliferative;  
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
KW central nervous system disorder; Alzheimer's disease; stroke;  
KW Parkinson's disease; Huntington's disease; coagulation disorder;  
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
KW tumour; infection; depression; psoriasis; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200021990-A1.  
PN  
XX 20-APR-2000.  
PD  
XX 15-OCT-1999; 99WO-US24205.  
PF  
XX 15-OCT-1998; 98US-0104435.  
PR  
XX (GEMY ) GENETICS INST INC.  
PA  
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M;  
PI WPI: 2000-317937/27.  
XX  
XX Isolated polynucleotides, and encoded proteins, comprising secreted  
PT expressed sequence tags (SESTs), useful for treating various disorders  
PT such as autoimmune, infectious, and central nervous system disorders -  
XX  
XX Claim 1; Page 354; 618pp; English.  
PS  
XX AAA41261 to AAA43419 represent specifically claimed secreted expressed  
CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat  
CC tissue sources. The SESTs can have a range of activities depending on  
CC the tissues they were isolated from. The activities include:  
CC chemokinetic; proliferative; immunomodulatory; haematopoietic;  
CC chemotactic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
CC antiasthmatic; vulnary; antitumor; osteopathic; neuroprotective;  
CC nootropic; antiparkinsonian; antiproliferative; cerebroprotective;  
CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
CC therapy and in vaccines. The SESTs are useful as probes for the  
CC identification and isolation of full-length cDNAs and genomic DNA  
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
CC are useful in assays for determining biological activity and raising  
CC antibodies. They may be useful for treatment of autoimmune disorders  
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
CC osteoporosis, osteoarthritis, central nervous system disorders  
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
CC disease), tumours, bacterial, fungal or viral infections, depression and  
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given  
CC in the exemplification of the present invention.  
XX  
SQ Sequence 241 BP; 59 A; 58 C; 58 G; 66 T; 0 other;  
Query Match 48.9%; Score 138; DB 21; Length 241;  
Best Local Similarity 100.0%; Pred. No. 9.1e-36;  
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 CTCGAGTTTCTGATCACAAGATCAGCGTGTGGTGGATCTCTCTAAACCATTTGTG 204  
Db 241 CTCGAGTTTCTGATCACAAGATCAGCGTGTGGTGGATCTCTCTAAACCATTTGTG 182





DT 26-SEP-2001 (first entry)  
XX Human protein tyrosine phosphatase (PTPase) DNA #2.  
XX Protein tyrosine phosphatase; PTPase; human; nervous system; dementia;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; autism;  
KW Tourette's syndrome; obsessive compulsive disorder; schizophrenia; mania;  
KW cardiovascular disorder; atherosclerosis; myocarditis; Addison's disease;  
KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; vitiligo;  
KW ankylosing spondylitis; Sjogren's syndrome; asthma; atopic dermatitis;  
KW diabetes mellitus; inflammatory disorder; pancreatitis; sarcoidosis; ss;  
KW allogeneic transplant rejection; multiple myeloma; Burkitt's lymphoma;  
KW leukaemia; cancer; neoplasm.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH CDS 230..1339  
FT /\*tag= a  
FT /product= "Human PTPase #2"  
XX WO200153530-A1.  
PN 26-JUL-2001.  
XX 17-JAN-2001; 2001WO-US01563.  
XX 18-JAN-2000; 2000US-0176306.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Shi Y, Ruben SM;  
XX WPI; 2001-442265/47.  
DR P-PSDB; AAU0486.  
XX New human protein tyrosine phosphatase polypeptide useful for treatment  
PT of disorders including disorder of the nervous system -  
XX Claim 1; Page 282-283; 290pp; English.  
XX The sequence represents DNA encoding a human protein tyrosine phosphatase  
CC (PTPase) polypeptide. PTPase proteins and nucleic acids are useful for  
CC preventing, treating or ameliorating a medical condition. The medical  
CC conditions are disorders of the central and peripheral nervous system  
CC (e.g. Alzheimer's disease, Huntington's disease, Parkinson's disease,  
CC Tourette's syndrome, obsessive compulsive disorder, schizophrenia, mania,  
CC dementia and autism), cardiovascular disorders (e.g. atherosclerosis and  
CC myocarditis), autoimmune disorders (e.g. rheumatoid arthritis, multiple  
CC sclerosis, ankylosing spondylitis, Addison's disease, Sjogren's syndrome,  
CC vitiligo, asthma, atopic dermatitis and diabetes mellitus), inflammatory  
CC disorders (e.g. pancreatitis, sarcoidosis and allogeneic transplant  
CC rejection) and cancers and neoplasms (e.g. leukaemia, Burkitt's lymphoma  
CC and multiple myeloma). The presence or absence of a mutation in the  
CC nucleic acid encoding the protein allows for diagnosis of or  
CC susceptibility to a pathological condition. The sequences are also useful  
CC in screening for agonists and antagonists of PTPase activity.  
XX Sequence 2872 BP; 840 A; 575 C; 601 G; 856 T; 0 other;  
SQ Query Match 23.1%; Score 65.2; DB 22; Length 2872;  
Best Local Similarity 54.1%; Pred. No. 5.1e-11;  
Matches 133; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
QY 35 TCGTGGATGACACCCAGGAGGCAATTGAAGTTCCACAAGAGCTCCTGGGAAGGTGCTGC 94  
DB 54 TGCTCGACGGGACCGAAGTGAGCGTGGACCTCGCGAAGACATGCCAAGGCCAGGATTGT 113  
QY 95 TGGATGCACTTTTGCAACACCTCAACCTCGTGGAGGTGACTATTGTCCTTCAGTTTC 154  
DB 114 TTGATCAGATTGTGTACCACTTGGACCTGTGTGGAACAGATTACTTTGGCCTCCAGTTCC 173  
QY 155 CTGATCACAAAAAGATCACGGTGTGGTGGATCTCTCTAAACCCATTGTGGAACAGATTA 214

Db 174 TCGACTCTGCCAGGTTGGCGCACTGGCTGGATCATGCCAAACCCATAAAAAAGCAGATGA 233  
QY 215 GAAGGCCAAACACACGTTGTTGTTAAAGTTTGGTGAATCTTTCCGGCTGACCACACAC 274  
Db 234 AAATTGGACCTGCTATGCTTTACACTTTTCGAGTTAAATACTATCTTCAGAACCAACA 293  
QY 275 AACATCC 280  
Db 294 ACCTTC 299  
RESULT 9  
ABN93400  
ID ABN93400 standard; cDNA; 3313 BP.  
XX AC ABN93400;  
XX 23-JUL-2002 (first entry)  
XX Human gene GS915528 coding sequence from chromosome 9q31-34 #2.  
XX Human; gene; chromosome 9q31-34; lipoprotein metabolism disorder;  
KW cholesterol transport disorder; ss.  
XX Homo sapiens.  
OS WO200071710-A2.  
PN 30-NOV-2000.  
PD 25-MAY-2000; 2000WO-FR01426.  
PF 25-MAY-1999; 99ER-0006587.  
PR 16-JUN-1999; 99US-0139450.  
XX (AVET ) AVENTIS PHARMA SA.  
PI Denefle P, Rosier-Montus M, Arnould-Reguigne I, Prades C, Clepet C;  
XX WPI; 2001-025161/03.  
DR P-PSDB; ABP37974.  
XX New nucleic acid derived from human chromosome 9, used e.g. for  
PT diagnosis and drug screening, derived from genes implicated in  
PT disorders of lipoprotein metabolism -  
XX Claim 1; Page 175-176; 269pp; French.  
XX The present sequence is the coding sequence for a human gene from  
CC chromosome 9q31-34. This sequence is likely to be involved in diseases  
CC of plasmatic lipoprotein metabolism, e.g. the reverse transport of  
CC cholesterol.  
XX Sequence 3313 BP; 922 A; 673 C; 719 G; 998 T; 1 other;  
SQ Query Match 23.1%; Score 65.2; DB 23; Length 3313;  
Best Local Similarity 54.1%; Pred. No. 5.5e-11;  
Matches 133; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
QY 35 TCGTGGATGACACCCAGGAGGCAATTGAAGTTCCACAAGAGCTCCTGGGAAGGTGCTGC 94  
Db 85 TGCTCGACGGGACCGAAGTGAGCGTGGACCTCGCGAAGACATGCCAAGGCCAGGATTGT 144  
QY 95 TGGATGCACTTTTGCAACACCTCAACCTCGTGGAGGTGACTATTGTCCTTCAGTTTC 154  
Db 145 TTGATCAGATTGTGTACCACTTGGACCTGTGTGGAACAGATTACTTTGGCCTCCAGTTCC 204  
QY 155 CTGATCACAAAAAGATCACGGTGTGGTGGATCTCTCTAAACCCATTGTGGAACAGATTA 214  
Db 205 TCGACTCTGCCAGGTGGCGCACTGGCTGGATCATGCCAACCCTATAAAAAAGCAGATGA 264  
QY 215 GAAGGCCAAACACACGTTGTGTGTTAAGTTGTGGTGAATCTTTCCGGCTGACCACACAC 274

```
Db 265 AAATTGGACCTGCTTATGCTTTTACACTTTTCGAGTTAAATACTATTCTTCAGAACCAACA 324
QY 275 AACTCC 280
Db 325 ACCTTC 330

RESULT 10
AAD32338
ID AAD32338 standard; DNA; 14917 BP.
XX
AC AAD32338;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human lung specific gene (LSG) 331878.
XX
KW Human; lung specific gene; LSG; lung embryonic development; cytostatic;
KW lung cancer; vaccine; gene therapy; non-cancerous lung disease; gene;
KW tumour; ds.
XX
OS Homo sapiens.
XX
PN WO200208278-A2.
XX
PD 31-JAN-2002.
XX
PF 20-JUL-2001; 2001WO-US22949.
XX
PR 21-JUL-2000; 2000US-219834P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Macina RA, Nair M, Chen S;
XX
DR WPI; 2002-268964/31.
XX
PT Novel lung specific gene useful for identifying, diagnosing,
PT monitoring, staging, imaging and treating lung cancer and non-cancerous
PT disease states in lung, for gene therapy, and for identifying lung
PT tissue.
XX
PS Claim 1; Page 131-136; 197pp; English.
XX
CC The present invention relates to lung specific genes (LSG) and their
CC corresponding polypeptides. LSG is useful for identifying, diagnosing,
CC monitoring, staging, imaging and treating lung cancer and non-cancerous
CC disease states in lung, identifying lung tissue, monitoring and
CC modifying lung embryonic development and differentiation, in gene
CC therapy, as hybridisation probes, to detect LSG mRNA as a marker for
CC lung cancer, as research reagents and materials for discovery of
CC treatments and diagnostics to human disease, to detect complementary
CC polynucleotides, and for chromosome identification. An antibody which
CC binds LSG is useful to detect or image localisation of LSG in a patient
CC for detecting or diagnosing a disease or condition, for preventing the
CC onset and treatment of lung cancer, to isolate or to identify clones
CC expressing LSG polypeptides, to purify LSG polypeptides, and to target
CC tumours expressing LSG. The present sequence is human LSG DNA.
XX
SQ Sequence 14917 BP; 3752 A; 3939 C; 3993 G; 3233 T; 0 other;

Query Match 21.8%; Score 61.4; DB 24; Length 14917;
Best Local Similarity 53.6%; Pred No. 2.1e-09;
Matches 128; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 35 TCCTGGATGACACCCAGGAGGCTTGAAGTTCCACAAAGAGCTCCTGGGAAGGTGCTGC 94
Db 8976 TGCTTGATGCTCGGAGTATGATGTGAGGTGGAGAAACATGCGCGGGCCAGGTGCTGT 9035
QY 95 TGGATGCAAGTTTGCACACCACTCAACCTCGTGGAGGTGACTATTTCGCCCTCGAGTTTC 154
Db 9036 TTGACCTGTGCTGTGAACACCTCAACCTCCTAGAGAAGGACTACTTCGGCCTGACCTTCT 9095
```

```
QY 155 CTGATCACAAAAGATCACGGTGTGGCTGGATCTCTCTAAACCCCATTTGTGAACAGATTA 214
Db 9096 GTGATGCTGACAGCCAGAGAAGAACTGGCTGGACCCCTCCAGGAGATCAAGAAGCAGATCC 9155
QY 215 GAAGGCCAAAGCACGTTTGTAGTTTGTGGTGAATTTCTTCGCCCTGACACACACA 273
Db 9156 GGAGTCAAGTGGCTTGTGTGTTGGGGAGGTGGTAGCCCTTGAATTTTGGCTTCACA 9214

RESULT 11
ABN93399
ID ABN93399 standard; cDNA; 3106 BP.
XX
AC ABN93399;
XX
DT 23-JUL-2002 (first entry)
XX
DE Human gene GS915528 coding sequence from chromosome 9q31-34 #1.
XX
KW Human; gene; chromosome 9q31-34; lipoprotein metabolism disorder;
KW cholesterol transport disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200071710-A2.
XX
PD 30-NOV-2000.
XX
PF 25-MAY-2000; 2000WO-FR01426.
XX
PR 25-MAY-1999; 99FR-0006587.
PR 16-JUN-1999; 99US-0139450.
XX
XX (AVET) AVENTIS PHARMA SA.
XX
PI Deneffe P, Rosier-Montus M, Arnould-Reguigne I, Prades C, Clepet C;
DR WPI; 2001-025161/03.
DR P-PSDB; ABP37973.
XX
PT New nucleic acid derived from human chromosome 9, used e.g. for
PT diagnosis and drug screening, derived from genes implicated in
PT disorders of lipoprotein metabolism.
XX
PS Claim 1; Page 174-175; 269pp; French.
XX
CC The present sequence is the coding sequence for a human gene from
CC chromosome 9q31-34. This sequence is likely to be involved in diseases
CC of plasmatic lipoprotein metabolism, e.g. the reverse transport of
CC cholesterol.
XX
SQ Sequence 3106 BP; 877 A; 603 C; 671 G; 955 T; 0 other;

Query Match 21.5%; Score 60.6; DB 23; Length 3106;
Best Local Similarity 53.6%; Pred No. 1.8e-09;
Matches 126; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 46 ACCCAGGAGGCATTTGAAGTTCACAAAGAGCTCTCTGGGAAGGTGCTGTGTGATGCAGTT 105
Db 1 ACCGAAGTGAGCGTGGACCTGCGCGAAACATGCCAAAGGCCAGGATTTGTTGATCAGATT 60
QY 106 TCAACACCCTCAACCTCGTGGAAAGTGACTATTTCGGCTCGAGTTTCCTGATCAAAA 165
Db 61 GTGTACCACCTTGAGCTCTGTGGAAACAGATTACTTTGGCCTCCAGTTCTCGACTCTGCC 120
QY 166 AAGATCACCGTGTGGCTGGATCTCTCTAAACCCATTGTCAAACAGATTAGAAGGCCAAG 225
Db 121 CAGGTGGCCACTGGCTGGATCATGCCAACCCATATAAAACAGATGAATTTGGACCT 180
QY 226 CAGGTGTTGTTAAGTTTGTGTGAAATTCCTTCGCCCTGACACACAACTCC 280
Db 181 GCTTATGCTTTTACACTTTTCGAGTTAAATACTATTCTTCAGAACCAACAACTTC 235
```









GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:33:49 ; Search time 8.57256 Seconds  
(without alignments)  
10088.332 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_151\_432  
Perfect score: 282  
Sequence: 1 ccttcaggaaactcgtgc.....ctgaccacacacaactccaa 282

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2.6/ptodata/1/ina/PTDUS\_COMB.seq: \*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	37.2	13.2	562	US-09-404-879A-147	Sequence 147, Appl
2	31.2	11.1	35100	US-08-770-379-18	Sequence 18, Appl
3	31.2	11.1	35100	US-08-757-669A-18	Sequence 18, Appl
4	31.2	11.1	35100	US-09-230-371A-18	Sequence 18, Appl
5	31	11.0	2680	US-09-063-035-1	Sequence 1, Appl
6	31	11.0	56516	US-08-996-306-1	Sequence 1, Appl
7	31	11.0	56516	US-09-338-907-1	Sequence 1, Appl
8	31	11.0	56516	US-09-218-207-1	Sequence 1, Appl
9	31	11.0	56520	US-09-338-907-179	Sequence 179, Appl
10	31	11.0	56520	US-09-218-207-179	Sequence 179, Appl
11	30.8	10.9	9775	US-09-647-540A-1	Sequence 1, Appl
12	30.8	10.9	9775	US-09-647-540A-3	Sequence 3, Appl
13	30.8	10.9	9775	US-09-647-540A-5	Sequence 5, Appl
14	30.8	10.9	9775	US-09-647-540A-7	Sequence 7, Appl
15	30.2	10.7	750	US-09-446-504-2	Sequence 2, Appl
16	30.2	10.7	750	US-09-712-266-2	Sequence 2, Appl
17	30.2	10.7	989	US-09-446-504-41	Sequence 41, Appl
18	30.2	10.7	989	US-09-712-266-41	Sequence 41, Appl
19	29.8	10.6	1082	US-09-484-970B-109	Sequence 109, Appl
20	29.4	10.4	12752	US-08-459-146-1	Sequence 1, Appl
21	29.4	10.4	12752	US-08-459-065-1	Sequence 1, Appl
22	29.2	10.4	1011	US-09-105-537-26	Sequence 26, Appl
23	29.2	10.4	1184	US-09-242-095-3	Sequence 3, Appl
24	29.2	10.4	36778	US-09-105-537-5	Sequence 5, Appl
25	28.4	10.1	2749	US-08-344-695-1	Sequence 1, Appl
26	28.4	10.1	8648	US-09-415-946-2	Sequence 2, Appl
27	28.4	10.1	17341	US-09-415-946-1	Sequence 1, Appl

c 28	28.2	10.0	3991	4	US-08-506-296B-3	Sequence 3, Appl
c 29	28.2	10.0	11958	4	US-09-134-246-8	Sequence 8, Appl
c 30	28	9.9	1698	4	US-09-059-769-20	Sequence 20, Appl
c 31	27.8	9.9	387	4	US-09-124-671-16	Sequence 16, Appl
c 32	27.8	9.9	515	3	US-08-758-662-5	Sequence 5, Appl
c 33	27.8	9.9	1537	4	US-09-556-877-57	Sequence 57, Appl
c 34	27.8	9.9	1537	4	US-09-620-412C-57	Sequence 57, Appl
c 35	27.8	9.9	2225	4	US-09-276-531-109	Sequence 109, Appl
c 36	27.8	9.9	3600	4	US-09-657-042A-3	Sequence 3, Appl
c 37	27.8	9.9	5010	3	US-09-161-244-1	Sequence 1, Appl
c 38	27.6	9.8	462	4	US-09-643-597-253	Sequence 253, Appl
c 39	27.6	9.8	604	4	US-09-123-912-90	Sequence 90, Appl
c 40	27.6	9.8	604	4	US-09-643-597-90	Sequence 90, Appl
c 41	27.6	9.8	2007	4	US-09-643-597-153	Sequence 153, Appl
c 42	27.6	9.8	3026	4	US-09-149-476-317	Sequence 317, Appl
c 43	27.6	9.8	3054	4	US-09-149-476-194	Sequence 194, Appl
c 44	27.4	9.7	612	2	US-08-308-952-13	Sequence 13, Appl
c 45	27.4	9.7	612	4	US-09-124-141-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1  
US-09-404-879A-147  
; Sequence 147, Application US/09404879A  
; Patent No. 6468546  
; GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.462C2  
; CURRENT APPLICATION NUMBER: US/09/404,879A  
; CURRENT FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 393  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 147  
; LENGTH: 562  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-404-879A-147

Query Match	13.2%	Score 37.2;	DB 4;	Length 562;
Best Local Similarity	53.4%;	Pred. No. 0.0022;		
Matches	78;	Conservative	0;	Mismatches 68; Indels 0; Gaps 0;
QY	43	GACACCCAGGAGGCATTTGAAGTTCACAAAGAGCTCCTGGGAAGGTGCTGGTGATGCA	102	
Db	175	GATGCAGAGCTGGAGTTTGCATCCAGCCAAATACAACTGGAAACAGCTTTTGTATCAG	234	
QY	103	GTTTGCAACACCTCAACTCTGTGGAGGTGACTATTTTGGCTCGAGTTTCTGTATCAC	162	
Db	235	GTGTGTAAGACTATCGGCTCCGGGAAGTGTGTACTTTTGGCTCCATATGTGGATAAT	294	
QY	163	AAAAAGATCACGGTGGCTGGATCT	188	
Db	295	AAAGGATTTCTACCTGGCTGAAGCT	320	

RESULT 2

US-08-770-379-18/C  
; Sequence 18, Application US/08770379  
; Patent No. 5849564  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF

NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/770,379  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 52342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35100 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-18

	Query Match	11.1%	Score 31.2;	DB 2;	Length 35100;
	Best Local Similarity	48.3%;	Pred. No. 3.2;		
	Matches 87;	Conservative 0;	Mismatches 93;	Indels 0;	Gaps 0;
Qy	103	GTTTTGAACACACCTCAACCTCGTGGAGGTGACATATTTTGGCCCTCGAGTTCTTCCTGATCAC	162		
Db	28764	GTTTTCAAGTGTCACCAACACGATACCAAGCAATTTTGACACAATTAACCTCAACTTGAA	28705		
Qy	163	AAAAAGATCAGCGTGTGGCTGGATCTCCTAAAAACCCATTGTGAAACAGATAGAAAGGCCA	222		
Db	28704	AAGGAGAGGAGCTCTGCATGAAGAGGCTAAATGTATTGAGACACACACTGTCACACCAA	28645		
Qy	223	AAGCACGTGTGTTTAACTTTGTGGTGAAATCTCTTCCGCTGACCCACACACAACCTCCAA	282		
Db	28644	CAGCCCGGTGACGCAAAAGGGTCCCGGGAGTGTGAATCTGCTGACCGCCAAACACATTTCAA	28585		

```

RESULT 3
US-08-757-669A-18/c
; Sequence 18, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
;   APPLICANT: Chang, Yuan
;   APPLICANT: Bohenzky, Roy A.
;   APPLICANT: Russo, James J.
;   APPLICANT: Edelman, Isidore S.
;   APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Cooper & Dunham LLP
;   STREET: 1185 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SS-08-757-669A-18

```

	Query Match	11.1%	Score 31.2;	DB 4;	Length 35100;
	Best Local Similarity	48.3%	Pred. No. 3.2;		
	Matches 87;	Conservative 0;	Mismatches 93;	Indels 0;	Gaps 0;
QY	103	GTTTGCACACCACTCGTGGAAAGTGACATATTTGGCCTCGAGTTCCCTGATCAC	162		
Db	28764	GTTTTCAAGTCTCTCACCAACGACAGATAACCAACAATTTGACACAAATAACTCAACTTGAA	28705		
QY	163	AAAAAGATCAGGTGTGGCTCGATCTCTAAAACCCATTGTGTAACAGATTAGAGGGCA	222		
Db	28704	AAGCAGAGGGAGCTCTGCATGAAGAGGCTAAATGTTATTTGAGACACAGCTGTCACACCA	28645		
QY	223	AAGCACGTTGTGTAAAGTTTGTGGTGAATTTCTTTCCGCCTGACACACACAACTCCAA	282		
Db	28644	CAGCCGGTGTACGCAAAAGGGTTCGCGGAGTGTGAATCTGTGACCGCCCAACACTTTCAA	28585		

```

RESULT 4
US-09-230-371A-18/c
; Sequence 18, Application US/092230371A
; Patent No. 6349586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIAT
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/23
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 35100
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associat
US-09-230-371A-18

```

	Query Match	11.18;	Score 31.2;	DB 4;	Length 35100;
	Best Local Similarity	48.3%;	Pred. No. 3.2;		
	Matches 87;	Conservative 0;	Mismatches 93;	Indels 0;	Gaps 0;
Qy 103	GTATTGCAACCACTCAACCTCGTGAAGGTGACTATTTTGGCGCTCGAGTTTCTCGATCAC	162			
Db 28764	GTTTTCAAGTGCTCCACCAACAGATAAACCAACAACTTGACACAACTAACTCAACTTGAA	28705			
Qy 163	AAAAAGATCAGGTGTGCTGGATCTCTCTAAACACCCATTGTGTAAACAGATTGAAGGCCA	222			

Db 28704 AAGGAGAGGAGCTGTCGATGAAGAGGCTAAATGTATTGAGACACAGCTGTACACCAA 28645  
QY 223 AAGCAGTGTGTTTAAAGTTTGGTGAATCTTTCCGCTGACACACACAACTCCAA 282  
Db 28644 CAGCCCGGTGACCAAGGTCGCCGGAGTGTGAATCTGCTACCGCCCAACACATTTCAA 28585

RESULT 5

US-09-063-035-1  
; Sequence 1, Application US/09063035  
; Patent No. 6160091  
; GENERAL INFORMATION:  
; APPLICANT: PEUKERT, Karen; HADNEL, Frank; and EILERS,  
; APPLICANT: Martin  
; TITLE OF INVENTION: Myc-binding zinc finger proteins,  
; TITLE OF INVENTION: their preparation and their use  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kell & Weinkauff  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage  
; COMPUTER: IBM AT-compatible, 80486 processor  
; OPERATING SYSTEM: MS-DOS version 6.1  
; SOFTWARE: WordPerfect version 8.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/063,035  
; FILING DATE: 21-APR-1998  
; CLASSIFICATION: 514  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2680 base pairs  
; TYPE: Nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 1..159  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 160..2571  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2572..2680  
US-09-063-035-1

Query Match 11.0%; Score 31; DB 4; Length 2680;  
Best Local Similarity 59.8%; Pred. No. 0.83;  
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 80 CTGGGAGGTGCTGGTGGATGAGTTGCAACCACTCAACCTCGTGGAGAGTGACTATT 139  
Db 2264 CTGTGAAGCAAGTGCAGGAAGAGCCCAACACTCACATCTCTACGCGCTGTGACTCCT 2323  
QY 140 TTGGCCTCGAGTTTCTGTATCAAAA 166  
Db 2324 GTGGGACAAGTTCTGGATGCAACA 2350

RESULT 6

US-08-996-306-1  
; Sequence 1, Application US/08996306  
; Patent No. 5945522  
; GENERAL INFORMATION:  
; APPLICANT: Cohen, Daniel

; APPLICANT: Chumakov, Ilya  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Bougueleret, Lydie  
; TITLE OF INVENTION: Prostate cancer gene  
; NUMBER OF SEQUENCES: 68  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 501 West Broadway  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-3505  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Win95  
; SOFTWARE: Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/996,306  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: GENSET.018A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 56516 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: DOUBLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: GENOMIC DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: Promoter  
; LOCATION: 1629..1870  
; IDENTIFICATION METHOD: Proscan  
; FEATURE:  
; NAME/KEY: Potential ATG  
; LOCATION: 1998..2000  
; FEATURE:  
; NAME/KEY: Exon 1  
; LOCATION: 2001..2216  
; FEATURE:  
; NAME/KEY: ATG  
; LOCATION: 2031..2033  
; FEATURE:  
; NAME/KEY: TYR phos  
; LOCATION: 11694..14332  
; FEATURE:  
; NAME/KEY: SEQ ID42  
; LOCATION: 11930..11947  
; FEATURE:  
; NAME/KEY: SEQ ID24  
; LOCATION: 12057..12103  
; FEATURE:  
; NAME/KEY: SEQ ID51  
; LOCATION: compl(12339..12358)  
; FEATURE:  
; NAME/KEY: SEQ ID64  
; LOCATION: 13547..13564  
; FEATURE:  
; NAME/KEY: SEQ ID58  
; LOCATION: 13657..13703  
; FEATURE:  
; NAME/KEY: SEQ ID67  
; LOCATION: compl(13962..13981)  
; FEATURE:  
; NAME/KEY: Exon 2

LOCATION: 18196..18265  
FEATURE: Exon 3  
NAME/KEY: 23717..23832  
LOCATION: Exon 4  
NAME/KEY: 25571..25660  
LOCATION: SEQ ID43  
NAME/KEY: 34216..34234  
LOCATION: FEATURE: SEQ ID25  
NAME/KEY: 34469..34515  
LOCATION: FEATURE: SEQ ID52  
NAME/KEY: compl(34625..34645)  
LOCATION: FEATURE: Exon 5  
NAME/KEY: 34669..34759  
LOCATION: FEATURE: Exon 6  
NAME/KEY: 40688..40846  
LOCATION: FEATURE: Exon 7  
NAME/KEY: 48070..48193  
LOCATION: FEATURE: Exon 8  
NAME/KEY: 50182..54523  
LOCATION: FEATURE: SEQ ID65  
NAME/KEY: 51149..51168  
LOCATION: FEATURE: SEQ ID59  
NAME/KEY: 51448..51494  
LOCATION: FEATURE: SEQ ID68  
NAME/KEY: compl(51482..51499)  
LOCATION: FEATURE: SEQ ID44  
NAME/KEY: 51596..51613  
LOCATION: FEATURE: SEQ ID26  
NAME/KEY: 51612..51658  
LOCATION: FEATURE: SEQ ID53  
NAME/KEY: compl(51996..52015)  
LOCATION: NAME/KEY: polyAd signal  
LOCATION: 54445..54450  
US-08-996-306-1

Query Match 11.0%; Score 31; DB 2; Length 56516;

Best Local Similarity 52.8%; Pred. No. 4.9;

Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 35 TGCTGATGACACCCAGGAGGCGATTGAAGTTCACAAAGAGCTCCTGGGAAGGTCGTC 94

Db 11827 TTCTGGAACACTGTGTAAGCTTTCAGTCCGAATAACATGATCATGCGCAAGTCTGT 11886

QY 95 TGGATGCACTTTGCAACCACTCAACCTCGTGGGAAGGTGACTATTTTGGCCCTCGAGTTTC 154

Db 11887 TAGATGTAGTCTGCAAGCATCTGTATTACTCTGGGCAAGACTATCTGTATTACAGCGG 11946

QY 155 CTGATCA 161

Db 11947 CTGATCA 11953

RESULT 7

US-09-338-907-1

; Sequence 1, Application US/09338907

; Patent No. 6265546

; GENERAL INFORMATION:

APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Ilya, Chumakov  
APPLICANT: Bougueleret, Lydie  
TITLE OF INVENTION: PROSTATE CANCER GENE  
FILE REFERENCE: GENSET.18CPICP  
CURRENT APPLICATION NUMBER: US/09/338,907  
EARLIER FILING DATE: 1999-06-23  
EARLIER APPLICATION NUMBER: 08/996,306  
EARLIER FILING DATE: 1997-12-22  
EARLIER APPLICATION NUMBER: 60/099,658  
EARLIER FILING DATE: 1998-09-09  
EARLIER APPLICATION NUMBER: 09/218,207  
EARLIER FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 578  
SOFTWARE: Patent.pm  
SEQ ID NO 1  
LENGTH: 56516  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: promoter  
LOCATION: 1629..1870  
OTHER INFORMATION: identification method Proscan  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1998..2000  
OTHER INFORMATION: potential start codon  
FEATURE:  
NAME/KEY: exon  
LOCATION: 2001..2216  
OTHER INFORMATION: exon1  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 2031..2033  
OTHER INFORMATION: ATG  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 11694..14332  
OTHER INFORMATION: Tyr Phos  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 11930..11947  
OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42  
FEATURE:  
NAME/KEY: allele  
LOCATION: 12057..12103  
OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 12339..12358  
OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complemen  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 13547..13564  
OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64  
FEATURE:  
NAME/KEY: allele  
LOCATION: 13657..13703  
OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 13962..13981  
OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complemen  
FEATURE:  
NAME/KEY: exon  
LOCATION: 18196..18265  
OTHER INFORMATION: exon 2  
FEATURE:  
NAME/KEY: exon  
LOCATION: 23717..23832  
OTHER INFORMATION: exon 3  
FEATURE:

[illegible]





OTHER INFORMATION: AATAAA	OTHER INFORMATION: AATAAA
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 1991..2008	LOCATION: 1991..2008
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 2305..2525	LOCATION: 2305..2525
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 4091..4111	LOCATION: 4091..4111
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 4528..4546	LOCATION: 4528..4546
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 5475..5495	LOCATION: 5475..5495
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 5927..5947	LOCATION: 5927..5947
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 8127..8144	LOCATION: 8127..8144
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 8560..8578	LOCATION: 8560..8578
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 11622..11639	LOCATION: 11622..11639
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 12018..12037	LOCATION: 12018..12037
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 11930..11947	LOCATION: 11930..11947
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 12339..12358	LOCATION: 12339..12358
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 12915..12932	LOCATION: 12915..12932
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 13317..13334	LOCATION: 13317..13334
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 13216..13233	LOCATION: 13216..13233
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 13617..13636	LOCATION: 13617..13636
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 13547..13564	LOCATION: 13547..13564
OTHER INFORMATION: upstream	OTHER INFORMATION: upstream
FEATURE:	FEATURE:
NAME/KEY: primer_bind	NAME/KEY: primer_bind
LOCATION: 13962..13981	LOCATION: 13962..13981
OTHER INFORMATION: downstream	OTHER INFORMATION: downstream

```

FEATURE:
NAME/KEY: primer_bind
LOCATION: 15994..16011
OTHER INFORMATION: downstream amplification primer 99-610
FEATURE:
NAME/KEY: primer_bind
LOCATION: 16463..16480
OTHER INFORMATION: upstream amplification primer 99-610 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17304..17324
OTHER INFORMATION: downstream amplification primer 99-609
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17814..17832
OTHER INFORMATION: upstream amplification primer 99-609 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18008..18027
OTHER INFORMATION: upstream amplification primer 4-90
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18423..18442
OTHER INFORMATION: downstream amplification primer 4-90 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18699..18716
OTHER INFORMATION: downstream amplification primer 99-607
FEATURE:
NAME/KEY: primer_bind
LOCATION: 19164..19182
OTHER INFORMATION: upstream amplification primer 99-607 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 22589..22609
OTHER INFORMATION: downstream amplification primer 99-602
FEATURE:
NAME/KEY: primer_bind
LOCATION: 23111..23129
OTHER INFORMATION: upstream amplification primer 99-602 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25098..25118
OTHER INFORMATION: downstream amplification primer 99-600
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25657..25674
OTHER INFORMATION: upstream amplification primer 99-600 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 26537..26557
OTHER INFORMATION: downstream amplification primer 99-598
FEATURE:
NAME/KEY: primer_bind
LOCATION: 27022..27040
OTHER INFORMATION: upstream amplification primer 99-598 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32822..32881
OTHER INFORMATION: downstream amplification primer 99-592
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32823..32841
OTHER INFORMATION: upstream amplification primer 99-592 , complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34215..34233
OTHER INFORMATION: upstream amplification primer 99-217
FEATURE:

```

Query Match 11.0%; Score 31; DB 4; Length 56520;  
Best Local Similarity 52.8%; Pred. NO. 4.9;  
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Query Match 11.0%; Score 31; DB 4; Length 56520;



```
NAME/KEY: primer_bind
LOCATION: 15994..16011
OTHER INFORMATION: downstream amplification primer 99-610
FEATURE:
NAME/KEY: primer_bind
LOCATION: 16463..16480
OTHER INFORMATION: upstream amplification primer 99-610, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17304..17324
OTHER INFORMATION: downstream amplification primer 99-609
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17814..17832
OTHER INFORMATION: upstream amplification primer 99-609, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18008..18027
OTHER INFORMATION: upstream amplification primer 4-90
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18423..18442
OTHER INFORMATION: downstream amplification primer 4-90, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18699..18716
OTHER INFORMATION: downstream amplification primer 99-607
FEATURE:
NAME/KEY: primer_bind
LOCATION: 19164..19182
OTHER INFORMATION: upstream amplification primer 99-607, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 22589..22609
OTHER INFORMATION: downstream amplification primer 99-602
FEATURE:
NAME/KEY: primer_bind
LOCATION: 23111..23129
OTHER INFORMATION: upstream amplification primer 99-602, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25098..25118
OTHER INFORMATION: downstream amplification primer 99-600
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25657..25674
OTHER INFORMATION: upstream amplification primer 99-600, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 26537..26557
OTHER INFORMATION: downstream amplification primer 99-598
FEATURE:
NAME/KEY: primer_bind
LOCATION: 27022..27040
OTHER INFORMATION: upstream amplification primer 99-598, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32262..32281
OTHER INFORMATION: downstream amplification primer 99-592
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32823..32841
OTHER INFORMATION: upstream amplification primer 99-592, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34215..34233
OTHER INFORMATION: upstream amplification primer 99-217
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34624..34644
```

Query Match 11.0%; Score 31; DB 4; Length 56520;  
Best Local Similarity 52.8%; Pred. No. 4.9;

```
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 35 TGTGTGATGACACCCAGGAGGCAATTTGAAGTTCCACAAAGAGCTCCTGGGAGGTCTGTC 94
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11827 TTTTGGAAAACACTGTGTAGCTTTTCAGTGGCAATAACATGATCAGTGGCAAGTTCTGT 11886
QY 95 TGGATCCAGTTTGGCAACCACTCAACCTCGTGGGAAGTGACTATTTTGGCTGCGAGTTTC 154
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11887 TAGATGTAGTCTGCAAGCATCTCTGATTCTTACTGGGCAAGACTATGTTGATTACAGCGG 11946
QY 155 CTGATCA 161
|||||
Db 11947 CTGATGA 11953
```

```
RESULT 11
US-09-647-540A-1/c
; Sequence 1, Application US/09647540A
; Patent No. 6420158
; GENERAL INFORMATION:
; APPLICANT: Ishii, Yoshitaka
; APPLICANT: Konishi, Jin
; APPLICANT: Hirasawa, Kazuaki
; APPLICANT: Okada, Hideki
; APPLICANT: Suzuki, Masanori
; TITLE OF INVENTION: GENES ENCODING DESULFURIZATION ENZYMES
; FILE REFERENCE: 11283-006001
; CURRENT APPLICATION NUMBER: US/09/647,540A
; PRIOR APPLICATION NUMBER: PCT/JP99/01756
; PRIOR FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: JP 310545/1998
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: JP 090387/1998
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 9775
; TYPE: DNA
; ORGANISM: Paenibacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3031)...(4410)
US-09-647-540A-1
```

Query Match 10.9%; Score 30.8; DB 4; Length 9775;  
Best Local Similarity 52.3%; Pred. No. 2.1;  
Matches 68; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
QY 127 GAAGGTGACTATTTTGGCCTCGAGTTTCCTGATCACAAGATCAGGTGGCTGGAT 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1164 GATTTTGGCGTGTGTTGGCGGCACTTCGGAGCAGTTTCAATTCAATGTTGTCGCCCTT 1105
QY 187 CTCCTAAACCCATTGTGAACAGATTAGAAGGCCAACAGCAGTTTCTGTTAAGTTTGTG 246
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1104 ATAAGCAATTCAGCGCGGATGCTCCGCATGTTCCAGCACATATTCCTTCAGTTCCGG 1045
QY 247 GTGAATTCCT 256
|||||
Db 1044 GTTAAACTCT 1035
```

```
RESULT 12
US-09-647-540A-3/c
; Sequence 3, Application US/09647540A
; Patent No. 6420158
; GENERAL INFORMATION:
; APPLICANT: Ishii, Yoshitaka
; APPLICANT: Konishi, Jin
; APPLICANT: Hirasawa, Kazuaki
; APPLICANT: Okada, Hideki
; APPLICANT: Suzuki, Masanori
```

APPLICANT: KATO, Ikunoshin

```
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/446,504
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 750
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
US-09-446-504-2

Query Match      10.7%; Score 30.2; DB 4; Length 750;
Best Local Similarity 50.3%; Pred. No. 0.75;
Matches 74; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 98 ATGCAGTTTGCACACCACTCGTGGAAGGTGACTATTTTGGCCTCGAGTTTCCTG 157
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 518 AGGAAGTTGAGATAAGCTAACTCTTGAAGATGAGGGATTATTCGACATCGAGGTTCAAG 577
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 158 ATCACAAAAGATCAGGGTGTGGCTGGATCTCCTAAACCCATTGTGAACACAGATTAGAA 217
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 578 AGGAGACAAAGAGCGCATATGGAGTCAGCTATCTCTCCGACATGGTTAAAGGACTTGGAA 637
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 218 GCCCAAGCACGTTGTGTTAAGTTTG 244
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 638 AGGCCGATGAAGTTACAATAAAGTTTG 664
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: December 3, 2002, 18:26:48  
Job time : 156.573 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 18:25:06 ; Search time 9.318 Seconds  
(without alignments)  
11655.134 Million cell updates/sec

Title: US-09-555-342b-1\_copy\_151\_432  
Perfect score: 282  
Sequence: 1 ccttcaggaaaactgtgtc.....ctgaccacacacaactccaa 282

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	71.6	25.4	466	10	US-09-728-445-750
2	65.2	23.1	2872	10	US-09-906-779-3
3	54.6	19.4	3984	10	US-09-848-294-1
4	45.2	16.0	4125	12	US-10-044-090-620
5	44.8	15.9	408	10	US-09-960-352-6773
6	44.2	15.7	417	12	US-10-044-090-624
7	38.8	13.8	300	10	US-09-998-598-1581
8	38.6	13.7	414	10	US-09-954-456-1461
9	37.2	13.2	512	10	US-09-960-253-27
10	37.2	13.2	562	10	US-09-884-441-147
11	37.2	13.2	2930	10	US-09-960-253-156
12	37.2	13.2	3044	10	US-09-880-107-3718
13	37.2	13.2	3047	10	US-09-864-864-329
14	37.2	13.2	3115	10	US-09-925-299-123
15	37	13.1	439	10	US-09-925-300-104
16	36.4	12.9	1372	10	US-09-765-111A-13
17	36.4	12.9	2334	10	US-09-765-111A-1
18	36.4	12.9	2523	10	US-09-765-111A-3
19	36.4	12.9	2596	10	US-09-765-111A-22

c 20	36.4	12.9	2625	10	US-09-765-111A-5	Sequence 5, Appli
c 21	36.4	12.9	2711	10	US-09-765-111A-24	Sequence 24, Appli
c 22	31	11.0	56516	9	US-09-853-526-1	Sequence 1, Appli
23	31	11.0	56516	10	US-09-901-484A-1	Sequence 1, Appli
24	31	11.0	56520	9	US-09-853-526-179	Sequence 179, App
25	31	11.0	56520	10	US-09-901-484A-179	Sequence 179, App
26	30.6	10.9	517	10	US-09-833-381-207	Sequence 207, App
27	30.6	10.9	517	10	US-09-833-381-208	Sequence 208, App
28	30.2	10.7	750	10	US-09-971-309-2	Sequence 2, Appli
29	30.2	10.7	989	10	US-09-971-309-41	Sequence 41, Appli
c 30	29.8	10.6	7574	10	US-09-070-927A-152	Sequence 152, App
c 31	29.6	10.5	695	10	US-09-867-701-10647	Sequence 10647, A
c 32	29.2	10.4	1011	9	US-09-860-846-26	Sequence 26, Appli
c 33	29.2	10.4	1011	10	US-09-861-289-26	Sequence 26, Appli
34	29.2	10.4	3957	12	US-10-005-467-1	Sequence 1, Appli
35	29.2	10.4	36778	9	US-09-860-846-5	Sequence 5, Appli
36	29.2	10.4	36778	10	US-09-861-289-5	Sequence 5, Appli
c 37	28.8	10.2	479	10	US-09-864-761-865	Sequence 865, App
c 38	28.8	10.2	4578	9	US-09-764-868-1401	Sequence 1401, Ap
39	28.8	10.2	4578	9	US-09-764-868-1402	Sequence 1402, Ap
40	28.6	10.1	392	10	US-09-833-381-1132	Sequence 1132, Ap
41	28.4	10.1	428	9	US-09-860-670-240	Sequence 240, App
c 42	28.4	10.1	790	10	US-09-880-107-416	Sequence 416, App
43	28.4	10.1	1693	10	US-09-822-849A-269	Sequence 269, App
44	28.4	10.1	4457	10	US-09-960-253-154	Sequence 154, App
45	28.4	10.1	9519	10	US-09-764-847-1056	Sequence 1056, Ap

ALIGNMENTS

RESULT 1  
US-09-728-445-750  
; Sequence 750, Application US/09728445  
; Patent No. US20020102543A1  
; GENERAL INFORMATION:  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and  
; TITLE OF INVENTION: Animals  
; FILE REFERENCE: LEX-0102-USA  
; CURRENT APPLICATION NUMBER: US/09/728,445  
; CURRENT FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/168,358  
; PRIOR FILING DATE: 1999-12-01  
; NUMBER OF SEQ ID NOS: 891  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 750  
; LENGTH: 466  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(466)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-728-445-750

Query Match 25.4%; Score 71.6; DB 10; Length 466;

Best Local Similarity 57.3%; Pred. No. 8.1e-15;

Matches 150; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

QY	24	CAAAATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTTCCACAAGAGCTCCTGG	83
DB	118	CAACATCTTCTCTGGATAACACTGTACAGGCTTTCAGAGTTTAAACAACATGATCAGG	177
QY	84	GAAGTGCCTGCTGGAGTCAGTTGCAACCACCTCAACCTCGTGGAGGTGACTATTTTGG	143
DB	178	GCAAGTCTGTTGGATATAGTCTTCAAGCATCTTGATTTGACTGACGAGACTATTTGG	237
QY	144	CCTCGAGTTTCCCT---GATCACAAAAGATCAGGGTGTGGCTGGATCTCCTAAAACCCAT	200
DB	238	TTTACAGTTGGTGACGATTCACAGATAACCCCAAGTGGCTGGATCCCAACAACCAAT	297





Db 515 AAAGCATCGAGTGCAGAAAGTACTCTCGATGGATCAGAAATATACCTGTGATGTAGAG 574

QY 70 CAAAGAGCTCCCTGGGAAGGTGCTGCTGGATGCAAGTTTGCAACACCACTCAACCTCGTGGAA 129

Db 575 AAAGCGTCCAGAGGACAAGTGTGTTGATAAAGTGTGGAACACTTGAACCTGTCTAGAG 634

QY 130 GGTGACTATTTTGGCTCGAGTTTCTGTATACACAAAAGATCAGCGTGTGGTGGATC 187

Db 635 AAAGACTACTTTGGGCTTACGTATCGAGATGCTGAAACCCAGAGAATTTGGTTGGACC 692

RESULT 5

US-09-960-352-6773

; Sequence 6773, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 6773

; LENGTH: 408

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 29-LIB34-005-Q1-E1-H1

US-09-960-352-6773

Query Match 15.9%; Score 44.8; DB 10; Length 408;

Best Local Similarity 51.2%; Pred. No. 1.1e-05;

Matches 131; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 29 TCCAGATGCTGGATGACACCCAGGAGGCAATTTGAAGTTCCACAAGAGCTCCTGGGAAGG 88

Db 111 TCCGCTTTTAGATGGCTTGGTACAGACCTTTAAAGTTAAACAACAAGATACCGGTCAAG 170

QY 89 TCGCTGCTGATCGAGTTTCCACACCACTCAACCTCGTGAAGGTGACTATTTTGGCCTCG 148

Db 171 TTCTCTGGACATGCTTACACTACCTGGGTGTGACTGAGAAGGAGTATTTTCGGTTTAC 230

QY 149 AGTTTCTCTGATCACAAAAGATCACGGTG---TGGCTGGATCTCTCTAAACCACTTGTGA 205

Db 231 ACATGGGTGATGACTCACTGGACTCTCTAGATGGCTTGAACCAAGCAACCACTCAGGA 290

QY 206 AACAGATTAGAAGGCCAAGACAGCTGTGTTAAGTTTGTGTAATTTCTTTCCGCCCTG 265

Db 291 AGCAGTTAAAGGAGGTTTCCCTCTACCTGATCCCTGATTTTCGAGTAAGATTTTATACCTG 350

QY 266 ACCACACAACTCCA 281

Db 351 ATCCCAACACACTGCA 366

RESULT 6

US-10-044-090-624

; Sequence 624, Application US/10044090

; Patent No. US20020137081A1

; GENERAL INFORMATION:

; APPLICANT: Olga Bandman

; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION

; FILE REFERENCE: PA-0028 US

; CURRENT APPLICATION NUMBER: US/10/044,090

; CURRENT FILING DATE: 2002-01-09

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: PERL Program

; SEQ ID NO 624

; LENGTH: 417

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; OTHER INFORMATION: Incyte ID No. US20020137081A1 347941.1

; NAME/KEY: unsure

; LOCATION: 355-356

; OTHER INFORMATION: a, t, c, g, or other

US-10-044-090-624

Query Match 15.7%; Score 44.2; DB 12; Length 417;

Best Local Similarity 57.7%; Pred. No. 1.8e-05;

Matches 79; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 82 GGAAGAGTCTGCTGTGATGACAGTTTGCAACCACTCAACCTCGTGAAGGTGACTATTTT 141

Db 96 GGACAAAGTCTGCTTTGATAAAGTGTGTGAACACTTGAACCTGCTAGAGAAAGACTACTT 155

QY 142 GGCCTCGAGTTTCCCTGATCACA AAAAGATCAGCGTGTGGCTGATCTCTAAACCCATT 201

Db 156 GGGCTTACGTATCGAGATGCTGAAAACCAAGAAATTTGGTTGGACCTCTCTAAGGAATA 215

QY 202 GTGAACACAGATTAGAG 218

Db 216 AAAAAACAGGTTGGAAG 232

RESULT 7

US-09-998-598-1581

; Sequence 1581, Application US/09998598

; Patent No. US20020150922A1

; GENERAL INFORMATION:

; APPLICANT: Stolk, John A.

; APPLICANT: Xu, Jiangchun

; APPLICANT: Chenault, Ruth A.

; APPLICANT: Meagher, Madelein Joy

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.561

; CURRENT APPLICATION NUMBER: US/09/998,598

; CURRENT FILING DATE: 2001-11-16

; NUMBER OF SEQ ID NOS: 2606

; SOFTWARE: Corixa Invention Disclosure Database

; SEQ ID NO 1581

; LENGTH: 300

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-998-598-1581

Query Match 13.8%; Score 38.8; DB 10; Length 300;

Best Local Similarity 56.2%; Pred. No. 0.0011;

Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 82 GGAAGAGTCTGCTGTGATGACAGTTTGCAACCACTCAACCTCGTGAAGGTGACTATTTT 141

Db 26 GGACAAAGTGTATTGACAAAGTGTGTGAACCACTCAATCTCTTGGAGAAAGACTACTT 85

QY 142 GGCCTCGAGTTTCCCTGATCACA AAAAGATCAGCGTGTGGCTGATCTCTAAACCCATT 201

Db 86 GGACTTTTGTTCAGGAAGCCCTGAGCAGAAAACCTGTTAGATCTCTCTAAGGAATA 145

QY 202 GTGAACACAGA 211

Db 146 AAGAGACAAA 155

RESULT 8

US-09-954-456-1461/c

; Sequence 1461, Application US/09954456

; Patent No. US20020115057A1

; GENERAL INFORMATION:

; APPLICANT: Young, Paul

; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C

; FILE REFERENCE: 1461/c

; CURRENT APPLICATION NUMBER: US/09/954,456

; CURRENT FILING DATE: 2002-01-09

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: PERL Program

; SEQ ID NO 1461

; LENGTH: 417

; TYPE: DNA



Matches	78;	Conservative	0;	Mismatches	68;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

Qy	43	GACACCCAGGAGCATTCTTGAAGTTCACAAAGAGCTCCTGGNAGGTGCTGCTGGATGCA	102
Db	113	GATGCAGAGCTGGAGTTTGCATTCACGCCAATACACACTGGAAACAGCTTTTGATCAG	172
Qy	103	GTTTGGACCACTCAACTCTGGTGAAGTGACTATTTTGGCCTCGAGTTTCTCTGATCAC	162
Db	173	GTGGTAAAGACTATFCGGCCTCGGGAAGTGTGGTACTTTGGCCTCCACTATGTGGATAAT	232
Qy	163	AAAAGATCACGGTGTGGCTGGATCT	188
Db	233	AAAGGATTTTCTTACCTGGCTGAAGCT	258

## RESULT 12

RESULTS  
US-09-880-107-3718

; Sequence 3718, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:

APPLICANT: Horne, Darci T.

: APPLICANT: Vockley, Joseph G.

APPLICANT: VOCKLEY, JOSEPH  
APPLICANT: scherf, ilse

APPLICANT: Scherr, Uwe

APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profile

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880,107

7 / CURRENT ATTENTION NUMBER: 03/03/000,10  
: CURRENT FILING DATE: 2001-06-14

/ CURRENT FILING DATE: 2001-06-14  
 : PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR APPLICATION NUMBER: US 60/211,379  
: PRIOR FILING DATE: 2000-06-14

; PRIOR FILING DATE: 2000-06-14  
 ; PRIOR REGISTRATION NUMBER: US 60/327 054

; PRIOR APPLICATION NUMBER: US 60/237,054

; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

SECRET ; SEQ ID NO 3718

: LENGTH: 3044

; LENGTH: 3044  
: TYPE: DNA

TYPE: DNA  
ORGANISM: Homo sapiens

ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No.

Query Match 13.28: Score 37 2: DB 10: Length 3044:

Query Match 13.2%, Score 37.2; DB 10; Length 3044;  
Best Local Similarity 53.4%; Pred. No. 0.013;

Best Local Similarity 33.4%, Fied: NO: 0.013,  
Matches 78: Conservative 0: Mismatches 68: Indels 0: Gaps 0:

[illegible]

RESULT. T 13

RESOL 13  
US-09-864-864-329

; Sequence 329, Application US/09864864

; Patent No. US20020102679A1

**GENERAL INFORMATION:**

APPLICANT: Xu, Jiangchun

APPLICANT: XU, JIANGCHUN  
APPLICANT: MITCHEM JENNIFER L

APPLICANT: MITCHELL, Jennifer L.

APPLICANT: Harlocker, Susan L.

APPLICANT: Dillon, Davin C.

; APPLICANT: Secrist, Heather

Qy 163 AAAAGATCACGGTGTGGCTGGATCT 188  
||| | | ||||| | |  
Db 306 AAAGGATTTCCCTACCTGGCTGAAGCT 331

## RESULT 15

```

US-09-925-300-104
; Sequence 104, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 104
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (360)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (402)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-104

```

Query Match 13.1%; Score 37; DB 10; Length 439;  
Best Local Similarity 52.2%; Pred. No. 0.0054;  
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY	24	CAAAATCCAGATGCTGGATGACACCCAGAGGCGATTGAAGTTCCACAAGAGCTCCTTGG	83
Db	169	CACATCCGGCTGCTGGACATCGGAGATCTCTGCCACATCCAGAGGGAACCAAGG	228
QY	84	GAAGTGCTGCTGGATGCGAGTTTGCAACCACTCAACCTCGTGGAGGTGACTATTTTGG	143
Db	229	GCAGTTTCTCATTGACCACATCTGCAACTACTACAGCCTCTGGAGAGGACTACTTTGG	288
QY	144	CTCTGAGTTTCTGTATCACAAAAGATCACGGTGTGG	180
Db	289	CATTGCGTATGTGGACCCAGAGAACCAAGGCATGG	325

Search completed: December 3, 2002, 22:25:28  
Job time : 16.318 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: December 3, 2002, 16:31:10 ; Search time 303.618 seconds  
(without alignments)  
15042.359 Million cell updates/sec

Title: US-09-555-342b-1\_COPY\_151\_432  
Perfect score: 282  
Sequence: 1 ccttcaggaactctgtc.....ctgaccacacacactccaa 282

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_htc:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vrt:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	282	100.0	859	9	AL537488
2	282	100.0	995	13	BM549907
3	278.8	98.9	450	9	AI690075
4	276	97.9	441	9	AI937921
5	274.4	97.3	433	9	AI298784
6	221.8	78.7	398	9	AI879532

c	7	215.4	76.4	482	12	BF011995
c	8	213.8	75.8	478	14	W82071
	9	212.2	75.2	629	12	BF021124
	10	208.2	73.8	472	12	BF940343
	11	207.4	73.5	1185	12	BG327626
c	12	193.6	68.7	354	12	B8835601
	13	178	63.1	631	12	BG019331
	14	177.2	62.8	1002	12	BG323892
	15	171	60.6	314	14	T05026
	16	170.4	60.4	612	14	BQ383992
	17	170.4	60.4	619	14	BQ397802
	18	156.4	55.5	298	12	B8816579
	19	150	53.2	570	9	AI091437
	20	132	46.8	570	12	BF195016
c	21	120	42.6	179	14	BQ305056
c	22	104.6	37.1	121	9	AI298786
	23	102.8	36.5	327	14	T07137
	24	97.8	34.7	679	12	BG772813
	25	97.8	34.7	752	13	BI763945
	26	97.8	34.7	788	12	BG718448
	27	97.8	34.7	971	14	BM910128
	28	97	34.4	741	9	AL555132
	29	92.2	32.7	517	9	AI528026
	30	92.2	32.7	626	13	BI697755
	31	92.2	32.7	635	10	BB654641
	32	92.2	32.7	652	10	BB663505
	33	92.2	32.7	898	14	BQ944256
	34	90.6	32.1	518	13	BI319861
	35	83.6	29.6	685	9	AUI68701
	36	82.8	29.4	488	10	BB840554
	37	81.8	29.0	325	12	BE816570
	38	72.8	25.8	416	12	BF914765
	39	72.4	25.7	350	13	BJ062492
	40	71.6	25.4	465	9	AA415407
	41	71.6	25.4	552	9	AI595558
	42	70.4	25.0	639	13	BM290978
	43	69	24.5	368	9	AA758137
	44	69	24.5	555	14	BM849444
	45	69	24.5	584	14	BM782015

ALIGNMENTS

RESULT 1  
AL537488 859 bp mRNA linear EST 13-FEB-2001  
LOCUS AL537488 LTI\_FLO13\_FBrnl Homo sapiens cDNA clone CS0DF025YJ01 5  
DEFINITION prime, mRNA sequence.  
ACCESSION AL537488  
VERSION AL537488.1 GI:12800981  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 859)  
AUTHORS Li W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
Bp 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
Location/Qualifiers  
1..859  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DF025YJ01"  
/clone\_lib="LTI\_FLO13\_FBrnl"  
/dev\_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"  
/lab\_host="DH10B"

/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 224 a 237 c 220 g 169 t 9 others  
ORIGIN

Query Match 100.0%; Score 282; DB 9; Length 859;  
Best Local Similarity 100.0%; Pred. No. 3.9e-81;  
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60  
DB 318 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 377  
QY 61 GAAGTTCCACAAGAGCTCCTGGAGGTGCTCGATGCTGGATGACACCCAGGAGCATTT 120  
DB 378 GAAGTTCCACAAGAGCTCCTGGAGGTGCTCGATGCTGGATGACACCCAGGAGCATTT 437  
QY 121 CTCGTGGAAGGTGACTATTTTGGCCTCGAGTTTCTCGATGACACCAAAAGATCAGGTGTGG 180  
DB 438 CTCGTGGAAGGTGACTATTTTGGCCTCGAGTTTCTCGATGACACCAAAAGATCAGGTGTGG 497  
QY 181 CTGGATCTCTAAACCCCATTTGGAACAGATTAGAGGCCAAGACACGTTTGTGTAAAG 240  
DB 498 CTGGATCTCTAAACCCCATTTGGAACAGATTAGAGGCCAAGACACGTTTGTGTAAAG 557  
QY 241 TTGTGTGTAATTTCTTCGGCTGACACACACACTCAA 282  
DB 558 TTGTGTGTAATTTCTTCGGCTGACACACACACTCAA 599

RESULT 2  
BM549907  
LOCUS BM549907 995 bp mRNA linear EST 20-FEB-2002  
DEFINITION AGENCOURT\_6558078 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5443230  
5', mRNA sequence.  
ACCESSION BM549907  
VERSION BM549907.1 GI:18785692  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 995)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone Distribution: Agencourt Bioscience Corporation  
found through the I.M.A.G.E. Consortium information can be  
http://image.llnl.gov  
Plate: LCM1920 row: g column: 07  
High quality sequence stop: 736.  
Location/Qualifiers  
1. .995  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5443230"  
/clone\_lib="NIH\_MGC\_98"  
/tissue\_type="astrocytoma grade IV, cell line"  
/lab\_host="DH10B (phage-resistant)"

FEATURES  
source  
Location/Qualifiers  
1. .995  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5443230"  
/clone\_lib="NIH\_MGC\_98"  
/tissue\_type="astrocytoma grade IV, cell line"  
/lab\_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRT; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 245 a 304 c 267 g 178 t 1 others  
ORIGIN

Query Match 100.0%; Score 282; DB 13; Length 995;  
Best Local Similarity 100.0%; Pred. No. 4.3e-81;  
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60  
DB 368 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 427  
QY 61 GAAGTTCCACAAGAGCTCCTGGAGGTGCTCGATGCTGGATGACACCCAGGAGCATTT 120  
DB 428 GAAGTTCCACAAGAGCTCCTGGAGGTGCTCGATGCTGGATGACACCCAGGAGCATTT 487  
QY 121 CTCGTGGAAGGTGACTATTTTGGCCTCGAGTTTCTCGATGACACCAAAAGATCAGGTGTGG 180  
DB 488 CTCGTGGAAGGTGACTATTTTGGCCTCGAGTTTCTCGATGACACCAAAAGATCAGGTGTGG 547  
QY 181 CTGGATCTCTAAACCCCATTTGGAACAGATTAGAGGCCAAGACACGTTTGTGTAAAG 240  
DB 548 CTGGATCTCTAAACCCCATTTGGAACAGATTAGAGGCCAAGACACGTTTGTGTAAAG 607  
QY 241 TTGTGTGTAATTTCTTCGGCTGACACACACTCAA 282  
DB 608 TTGTGTGTAATTTCTTCGGCTGACACACACTCAA 649

RESULT 3  
AI690075/c  
LOCUS AI690075 450 bp mRNA linear EST 16-DEC-1999  
DEFINITION tx27d07.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2270797 3', similar to WP:H05G16.1 CELL1568 BAND 4.1 FAMILY ;, mRNA sequence.  
ACCESSION AI690075  
VERSION AI690075.1 GI:4901369  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 450)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone Distribution: Washington University Genome Sequencing Center  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 673 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 444.  
Location/Qualifiers  
1. .450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2270797"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"

FEATURES  
source  
Location/Qualifiers  
1. .450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2270797"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"

/lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; plasmid DNA from the normalized library NCI CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo. "  
 BASE COUNT 106 a 111 c 109 g 124 t  
 ORIGIN

Query Match 98.9%; Score 278.8; DB 9; Length 450;  
 Best Local Similarity 99.3%; Pred. No. 3e-80;  
 Matches 280; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CCTTCAGGAAGACTCTGTCCATCAAAATCCAGATCGTGGATGACACCCAGGAGGCATT 60  
 Db 439 CCTCAAGGAAACTCTGTCCATCAAAATCCAGATCGTGGATGACACCCAGGAGGCATT 380  
 QY 61 GAAGTTCACAAAGAGCTCTCGTGGAGGTGCTGCTGGATGCAGTTTGCACACCACTCAAC 120  
 Db 379 GAAGTTCACAAAGAGCTCTCGTGGAGGTGCTGCTGGATGCAGTTTGCACACCACTCAAC 320  
 QY 121 CTGCTGGAAGGTGACTATTGTCGCCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTGG 180  
 Db 319 CTGCTGGAAGGTGACTATTGTCGCCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTGG 260  
 QY 181 CTGGATCTCTAAACCCATTGTCGAGTTTCCTGATCACAAAAGATCACGGTGTGG 240  
 Db 259 CTGGATCTCTAAACCCATTGTCGAGTTTCCTGATCACAAAAGATCACGGTGTGG 200  
 QY 241 TTTGTGCTGAAATCTTTCGCGCTGACACACACCACTCAA 282  
 Db 199 TTTGTGCTGAAATCTTTCGCGCTGACACACCACTCAA 158

RESULT 4  
 AI937921/c  
 LOCUS  
 DEFINITION  
 au54f07.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
 IMAGE:2518597 3' similar to WP:H05G16.1 CELL1568 BAND 4.1 FAMILY ;,  
 mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Kitzman, D., Kucaba, T., Lacy, M., Lennon, G., Marra, M., Martin  
 , J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
 White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 Unpublished (1997)

JOURNAL  
 COMMENT  
 Contact: wilson rk  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 429.

FEATURES  
 source

1..441  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

/clone="IMAGE:2518597"  
 /clone\_lib="Schneider fetal brain 00004"  
 /sex="male"  
 /tissue\_type="frontal lobe"  
 /dev\_stage="5 months post-conception"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pBluescript SK (Stratagene);  
 Site\_1: SstI; Site\_2: XhoI; Double-stranded cDNA was  
 prepared from human fetal brain tissue. 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence:  
 5'-GAGAGAGAGAGCTCAAGATCCTTAATTAATATATCCCCCCCCC-3'  
 and 3' adaptor sequence:  
 5'-GAGAGAGAGACTCGAGTTTTTTTTTTT-3'. The library was  
 size-selected for >0.5 kb inserts and has an average  
 insert size estimated at 1.2 kb. This library was  
 constructed using the CAP-trapper method for full-length  
 enrichment and has not undergone amplification. Library  
 was constructed by Dr. Claudio Schneider (LNCIB-Area  
 Science Park, Trieste, Italy)."  
 BASE COUNT 105 a 110 c 100 g 126 t  
 ORIGIN

Query Match 97.9%; Score 276; DB 9; Length 441;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-79;  
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 GGAAGAACTCGTCCATCAAAATCCAGATCCTGGATGACACCCAGGAGGCATTGAAGTT 66  
 Db 441 GGAAGAACTCGTCCATCAAAATCCAGATCCTGGATGACACCCAGGAGGCATTGAAGTT 382  
 QY 67 CCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGACACCACTCAACCTCGTG 126  
 Db 381 CCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGACACCACTCAACCTCGTG 322  
 QY 127 GAAGGTGACTATTGTCGCTCGAGTTTCCTGATCACAAAAGATCAGGTGTGCTGGAT 186  
 Db 321 GAAGGTGACTATTGTCGCTCGAGTTTCCTGATCACAAAAGATCAGGTGTGCTGGAT 262  
 QY 187 CTCCTAAACCCATTGTGAACAGATTAGAGCCCAAGCAGCTTTGTTAAGTTTG 246  
 Db 261 CTCCTAAACCCATTGTGAACAGATTAGAGCCCAAGCAGCTTTGTTAAGTTTG 202  
 QY 247 GTGAAATCTTTCGCGCTGACACACACCACTCAA 282  
 Db 201 GTGAAATCTTTCGCGCTGACACACACCACTCAA 166

RESULT 5  
 AI298784/c  
 LOCUS  
 DEFINITION  
 qm91b09.x1 NCI CGAP Lu5 Homo sapiens cDNA clone IMAGE:1896089 3'  
 similar to WP:H05G16.1 CELL1568 ;, mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 433)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be



found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bbrp/image/image.html  
 Insert length: 702 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 359.

# FEATURES

Location/Qualifiers  
 1. .433  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1896089"  
 /clone\_lib="NCI\_CGAP\_Lu5"  
 /tissue\_type="carcinoid"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from a neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 105 a 111 c 98 g 119 t  
 ORIGIN  
 Query Match 97.3%; Score 274.4; DB 9; Length 433;  
 Best Local Similarity 99.6%; Pred. No. 8.2e-79;  
 Matches 275; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 7 GGAAACTCGTGCCATCAAAATCCAGATGCTGGATGACACCCAGAGGCATTTGAAGTT 66  
 Db 433 GGAAACTCGTGCCATCAAAATCCAGATGCTGGATGACACCCAGAGGCATTTGAAGTT 374  
 QY 67 CCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGAGTTTGCACCACTCAACCTCGTG 126  
 Db 373 CCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGAGTTTGCACCACTCAACCTCGTG 314  
 QY 127 GAAGTGACTATTTTGGCCCTCGAGTTTCTGTATCAGAAAAGATCACGGTGTGGTGGAT 186  
 Db 313 GAAGTGACTATTTTGGCCCTCGAGTTTCTGTATCAGAAAAGATCACGGTGTGGTGGAT 254  
 QY 187 CTCCTAAACCCATTTGTAACAGATTAGAAGCCAAACAGCTGTTGTTTGAAGTTTGTG 246  
 Db 253 CTCCTAAACCCATTTGTAACAGATTAGAAGCCAAACAGCTGTTGTTTGAAGTTTGTG 194  
 QY 247 GTGAATTTCTTCCGCTGACACACACACAACTCCAA 282  
 Db 193 GTGAATTTCTTCCGCTGACACACACACAACTCCAA 158

# RESULT 6

AI879532/c  
 LOCUS AI879532 398 bp mRNA linear EST 23-AUG-1999  
 DEFINITION au54c11.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone  
 IMAGE:2518580 3' similar to WP:H05G16.1 CE11568 BAND 4.1 FAMILY ;,  
 mRNA sequence.

ACCESSION AI879532  
 VERSION AI879532.1 GI:55533590  
 KEYWORDS EST.  
 SOURCE human.

# ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

# REFERENCE

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin  
 ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
 White,X., Wylie,T., Waterston,R. and Wilson,R.

# TITLE

WashU-NCI human EST Project

# JOURNAL

Unpublished (1997)

# COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 245.  
 Location/Qualifiers  
 1. .398  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2518580"  
 /clone\_lib="Schneider fetal brain 00004"  
 /sex="male"  
 /tissue\_type="frontal lobe"  
 /dev\_stage="5 months post-conception"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pBluescript SK (Stratagene);  
 Site.1: SstI; Site.2: XhoI; Double-stranded cDNA was  
 prepared from human fetal brain tissue. 5' and 3'  
 adaptors were used in cloning as follows: 5' adaptor  
 sequence:  
 5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCCCCCCCCC-3'  
 and 3' adaptor sequence:  
 5'-GAGAGAGAGAGCTCGAGTTT-3'. The library was  
 size-selected for >0.5 kb inserts and has an average  
 insert size estimated at 1.2 kb. This library was  
 constructed using the CAP-trapper method for full-length  
 enrichment and has not undergone amplification. Library  
 was constructed by Dr. Claudio Schneider (LNCIB-Area  
 Science Park, Trieste, Italy)."

BASE COUNT 99 a 98 c 86 g 115 t  
 ORIGIN  
 Query Match 78.7%; Score 221.8; DB 9; Length 398;  
 Best Local Similarity 99.1%; Pred. No. 1.3e-61;  
 Matches 223; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 58 TTGAGTTCCACAAAGAGCTCTGGGAGGTGCTGCTGGATGACAGTTTGCACCACTC 117  
 Db 398 TTGAGTTCCACAGAGAGCTCTGGGAGGTGCTGCTGGATGACAGTTTGCACCACTC 339  
 QY 118 AACCTCGTGAAGGTGACTATTTTGGCCCTCGAGTTTCTGTATCAGAAAAGATCAGGTG 177  
 Db 338 AACCTCGTGAAGGTGACTATTTTGGCCCTCGAGTTTCTGTATCAGAAAAGATCAGGTG 279  
 QY 178 TGGCTGGATCTCCTAAACCCATTTGTAACAGATTAGAAGCCAAACAGCTGTTGTTT 237  
 Db 278 TGGCTGGATCTCCTAAACCCATTTGTAACAGATTAGAAGCCAAACAGCTGTTGTTT 219  
 QY 238 AAGTTGTGTCGAATCTTCCGCTGACACACACAACTCCAA 282  
 Db 218 AAGTTGTGTCGAATCTTCCGCTGACACACACAACTCCAA 174

# RESULT 7

BF011995/c

LOCUS BF011995 482 bp mRNA linear EST 06-OCT-2000

DEFINITION ux32g02.y1 Soares\_NKMD\_mandible Mus musculus cDNA clone

IMAGE:3513938 5' similar to TR:Q9Y4F1 Q9Y4F1 CDEP. ;, mRNA

sequence.

ACCESSION BF011995

VERSION BF011995.1 GI:10712270

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 482)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)



```

ACCESSION   BF021124
VERSION     BF021124.1  GI:10752456
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 629)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Other_ESTs: ux52g02.y1
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1396818
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 429.
            Location/Qualifiers
                1..629
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="IMAGE:3513938"
                /clone_lib="Soares_NKWD_mandible"
                /tissue_type="mandible"
                /lab_host="DH10B (phage-resistant)"
                /note="Vector: pT73D-Pac (Pharmacia) with a modified
                polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
                was primed with a Not I - oligo(dT) primer [5'
                TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTGTGTTT 3'],
                double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT73 vector. Library
                went through one round of normalization. Library
                constructed by Bento Soares and M. Fatima Bonaldo. "
            BASE COUNT      140 a 205 c 171 g 113 t
            ORIGIN
            Query Match      75.2%; Score 212.2; DB 12; Length 629;
            Best Local Similarity 84.7%; Pred. No. 2.4e-58;
            Matches 238; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATCGTGATGACACACCCAGGAGCATTT 60
    |||||
Db 258 CCTTCAGGAAACTTATGACTGTCAAAATTCAGATCGTGATGACACACCCAGGAGCATTT 317
    |||||

QY 61 GAAGTTCCACAAAGAGCTCGTGGAAAGGTGCTGCTGGATGCAGTTTGCACACCACTCAAC 120
    |||||
Db 318 GAAGTTCCACAAAGAGCGCCAGGAGGTCTGTTTGACGCCGTGTGCACACCACTCAAC 377
    |||||

QY 121 CTGCTGGAAGGTGACTATTTTGGCCCTCGAGATTTCTGTGATCACAACCAAGATCAGGTTGG 180
    |||||
Db 378 CTGCTGGAAGGTGACTACTTCGGCCCTGGAGTTCCCTGACCAAGGAAGATCGTGGTGG 437
    |||||

QY 181 CTGATCTCTTAAACCCATTGTGAACACAGATTAGAGGCCAAGCACAGTCTGTTGTTAAG 240
    |||||
Db 438 CTGATCTCTTGAAGCCCATTTGTGAAGCAGATTTCGAAGCCGAGGACCGTGTGTTTAA 497
    |||||

QY 241 TTTGTGGTGAATTTCTTCCGCTCGACCAACCACTCA 281
    |||||
Db 498 TTTGTGTGAATTTCTTCCCTCCGACCAACCACTGCA 538
    |||||

RESULT 10
BF940343
LOCUS      nad95e02.x1 NCI-CGAP_Ov18 Homo sapiens cDNA clone IMAGE:343343 3'
DEFINITION similar to TR:Q9Y4F1 Q9Y4F1 CDEP.;, mRNA sequence.
ACCESSION BF940343
VERSION   BF940343.1  GI:12357663
KEYWORDS  EST.

SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 472)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -400P from Gibco
            High quality sequence stop: 378.
            Location/Qualifiers
                1..472
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3433443"
                /clone_lib="NCI-CGAP_Ov18"
                /tissue_type="fibrotheoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a
                modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                strand cDNA was primed with a Not I - oligo(dT) primer [5'
                TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTGTGTTT 3'],
                double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not
                I and Eco RI sites of the modified pT73 vector. Library
                went through one round of normalization, and was
                constructed by Bento Soares and M. Fatima Bonaldo. "
            BASE COUNT      105 a 145 c 142 g 80 t
            ORIGIN
            Query Match      73.8%; Score 208.2; DB 12; Length 472;
            Best Local Similarity 98.6%; Pred. No. 4.2e-57;
            Matches 210; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATCGTGATGACACCCAGGAGCATTT 60
    |||||
Db 260 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATCGTGATGACACCCAGGAGCATTT 319
    |||||

QY 61 GAAGTTCCACAAAGAGCTCGTGGAAAGGTGCTGCTGGATGCAGTTTGCACACCACTCAAC 120
    |||||
Db 320 GAAGTTCCACAAAGAGCTCGTGGAAAGGTGCTGCTGGATGCAGTTTGCACACCACTCAAC 379
    |||||

QY 121 CTGCTGGAAGGTGACTATTTTGGCCCTCGAGTTTCTGTGATCACAACCAAGATCAGGTTGG 180
    |||||
Db 380 CTGCTGGAAGGTGACTATTTTGGCCCTCGAGTTTCTGTGATCACAACCAAGATCAGGTTGG 439
    |||||

QY 181 CTGGATCTCTTAAACCCCATTTGTGAACACAGATT 213
    |||||
Db 440 GTGGATCTTCTCAACCCCATTTGTGAACACAGATT 472
    |||||

RESULT 11
BG327626
LOCUS      602426620F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564636 5',
DEFINITION mRNA sequence.
ACCESSION BG327626
VERSION   BG327626.1  GI:13134064
KEYWORDS  EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 1185)  
TITLE NIH-MGC http://mgc.nci.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCMI280 row: 9 column: 05  
High quality sequence stop: 564.

FEATURES  
Location/Qualifiers

1. .1185  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4564636"  
/clone\_lib="NIH\_MGC\_14"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pORF7; Site\_1: XhoI; Site\_2:  
EcoRI; CDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adapter: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 378 a 377 c 271 g 159 t

Query Match 73.5%; Score 207.4; DB 12; Length 1185;  
Best Local Similarity 94.2%; Pred. No. 1.3e-56;  
Matches 259; Conservative 0; Mismatches 11; Indels 5; Gaps 4;  
QY 13 CTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTTCCACAA 72  
|||||  
Db 2 CTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTTCCACAA 61  
QY 73 AGAGCTCCTGGGAAGTGTCTGGATGACAGTTTGAACACAC-CTCAACCTCGTGGGAAG 131  
|||||  
Db 62 AGAGCTCCTGGGAAGTGTCTGGATGACAGTTTGAACACACCTCAACCTCGTGGACGG 121  
QY 132 TGACTATTTT-GGCTTCGAGTTTCTTGATCACAAAGATCACGGTGTGGCTGATCTCC 190  
|||||  
Db 122 TGACTATTTTGGCCCTCGAGTTTCTTGATCACAAAGATCACGGTGTGGCTGATCTCC 181  
QY 191 TAAACCCATTGTGAA-ACAGATTAGAAGGCCAAAGACAGTTGTT--GTTAAGTTTGTGG 247  
|||||  
Db 182 TAGAACCCATTGTGAACACAGATTAGAAGGCCAAAGACAGTTTGTGTTAAGTTTGTGG 241  
QY 248 TGAATTCCTTCGCGCTGACCAACACACTCCAA 282  
|||||  
Db 242 TGAATTCCTTCGCGCTGACCAACACACTCCAA 276

RESULT 12  
LOCUS BE835601/c 354 bp mRNA linear EST 22-SEP-2000  
DEFINITION RC6-FN0027-110700-011-F03 FN0027 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE835601  
VERSION BE835601.1 GI:10267979  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 354)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

TITLE JOURNAL  
MEDLINE  
COMMENT

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC6-FN0027-110  
700-011-F03&t3=2000-07-11&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 43  
High quality sequence stop: 353.

FEATURES  
Location/Qualifiers

1. .354  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="FN0027"  
/dev\_stage="Adult"  
/note="Organ: prostate\_normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the pUC 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."  
BASE COUNT 98 a 89 c 81 g 86 t

Query Match 68.7%; Score 193.6; DB 12; Length 354;  
Best Local Similarity 98.0%; Pred. No. 2.2e-52;  
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 83 GGAAGTGCTGCTGGATGTCAGTTTGCACCCACCTCAACCTCGTGGAGGTGACTATTTTG 142  
|||||  
Db 354 GGATGCTGCTGGATGTCAGTTTGCACCCACCTCAACCTCGTGGAGGTGACTATTTTG 295  
QY 143 GCCTCAGTTTCTGATCACAAAAGATCACGGTGTGGTGCATCTCTAAAACCCATTG 202  
|||||  
Db 294 GCCTCAGTTTCTGATCACAAAAGATCACGGTGTGGTGCATCTCTAATACCCATTG 235  
QY 203 TGAACAGATTAGAAGGCCAAAGCAGTTGTTCTTAAAGTTTGGTGAATCTTTCCCG 262  
|||||  
Db 234 TGAACAGATTAGAAGGCCAAAGCAGTTGTTCTTAAAGTTTGGTGAATCTTTCCCG 175  
QY 263 CTGACCACACACAACTCCAA 282  
|||||  
Db 174 CTGACCACACAACTCCAA 155

RESULT 13  
LOCUS BG019331  
DEFINITION BG019331 631 bp mRNA linear EST 24-JAN-2001  
IMAGE:4083497 5' similar to TR:Q914F1 Q914F1 CDEP. ;, mRNA  
sequence.  
ACCESSION BG019331  
VERSION BG019331.1 GI:12475410  
KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 314)  
TITLE Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.  
3,400 expressed sequence tags identify diversity of transcripts  
from human brain  
JOURNAL Nat. Genet. 4, 256-267 (1993)  
MEDLINE 93364420  
COMMENT Contact: Adams, MD  
The Institute for Genomic Research  
932 Clopper Road, Gaithersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423  
Email: mdadams@tigr.org  
Seq primer: M13-21.

FEATURES  
source  
1..314  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):81769"  
/db\_xref="taxon:9606"  
/clone="HF8CP32"  
/clone\_lib="Fetal brain, Stratagene (cat#936206)"  
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female;  
oligo-dt + random primed cDNA synthesis; lambdaZAP-II  
vector, 1.0kb average inser size."  
BASE COUNT 82 a 78 c 87 g 63 t 4 others  
ORIGIN

Query Match 60.6%; Score 171; DB 14; Length 314;  
Best Local Similarity 98.4%; Pred. No. 5.1e-45;  
Matches 182; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CCTTCAGGAAACTCGTGCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60  
Db |||||  
130 CCTTCAGGAAACTCGTGCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 189  
QY 61 GAAGTTCCACAAGAGCTCCTGGGAAGTGCTGCTGGATGCAAGTTGCAACCACTCAAC 120  
Db |||||  
190 GAAGTTCCACAAGAGCTCCTGGGAAGTGCTGCTGGATGCAAGTTGCAACCACTCAAC 249  
QY 121 CTCGT-GGAAGTGACTATTTGGCTCGAGTTTCCTGATCACAAAAAGATCACGCTGTG 179  
Db |||||  
250 CTCGTGGGAAGGTGACTATTTGGNCTCGAGTTTCTGATCACAAAAAGATCACGCTGTG 309  
QY 180 GCTGG 184  
Db |||||  
310 GCTGG 314

Search completed: December 3, 2002, 21:24:16  
Job time : 318.618 secs

**THIS PAGE BLANK (USPTO)**